

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
31 July 2003 (31.07.2003)

PCT

(10) International Publication Number  
**WO 03/061681 A2**

(51) International Patent Classification<sup>7</sup>: **A61K 38/00**,  
39/395, 48/00, 31/7088, C12N 15/12, C07K 14/47, 16/18,  
A01K 67/00, C12Q 1/68, G01N 33/50, A61K 45/00,  
A61P 3/00

[DE/DE]; Am Graben 10b, 37079 Göttingen (DE).  
**BRÖNNER, Günter** [DE/DE]; Springstrasse 54, 37077  
Göttingen (DE).

(21) International Application Number: PCT/EP03/00738

(74) Agents: **WEICKMANN & WEICKMANN** et al.; Post-  
fach 860 820, 81635 München (DE).

(22) International Filing Date: 24 January 2003 (24.01.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
02 001 806.5 25 January 2002 (25.01.2002) EP  
02 003 473.2 14 February 2002 (14.02.2002) EP  
02 004 687.6 28 February 2002 (28.02.2002) EP  
02 009 475.1 25 April 2002 (25.04.2002) EP  
02 013 329.4 18 June 2002 (18.06.2002) EP  
02 029 081.3 30 December 2002 (30.12.2002) EP

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GI,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE,  
SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,  
VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,  
ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI,  
SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN,  
GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant (*for all designated States except US*): **DEVELO-  
GEN AKTIENGESELLSCHAFT FÜR ENTWICK-  
LUNGSBIOLOGISCHE FORSCHUNG** [DE/DE];  
Rudolf-Wissell-Strasse 28, 37079 Göttingen (DE).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **STEUERNAGEL,  
Arnd** [DE/DE]; Am Kirschberge 4, 37085 Göttingen  
(DE). **MOLITOR, Andreas** [DE/DE]; Max-Both-Ring  
2F, 37077 Göttingen (DE). **EULENBERG, Karsten**

**Published:**

— *without international search report and to be republished  
upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.*

(54) Title: PROTEINS INVOLVED IN THE REGULATION OF ENERGY HOMEOSTASIS AND ORGANELLE METABOLISM

(57) Abstract: This invention relates to the use of nucleic acid and amino acid sequences of Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta and epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, human KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, human KIAA0095 protein, formin-binding protein 21, and/or homologous proteins in pharmaceutical compositions, and to the use of these sequences in the diagnosis, study, prevention, and treatment of diseases and disorders related to body-weight regulation and thermogenesis.



WO 03/061681 A2

## Proteins involved in the regulation of energy homeostasis and organelle metabolism

5

### Description

This invention relates to the use of nucleic acid and amino acid sequences of Optic atrophy 1 protein (OPA1), cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or formin-binding protein 21, or a homologous protein in pharmaceutical compositions, and to the use of these sequences and to the use of effectors thereof in the diagnosis, study, prevention, and treatment of diseases and disorders related to body-weight regulation and thermogenesis, for example, but not limited to, metabolic diseases such as obesity, as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis and gallstones, and disorders related to ROS defence, such as diabetes mellitus and neurodegenerative disorders.

Mitochondria are the energy suppliers of animal cells. Most of the energy available from metabolising foodstuffs like carbohydrates, fats etc. is used to create a proton gradient across the inner mitochondrial membrane. This proton gradient drives the enzyme ATP synthetase that produces ATP, the cells major fuel substance (Mitchell P, Science 206, 1979, 1148-1159). In the mitochondria of brown adipose tissue exists a protein (Uncoupling Protein 1) that tunnels protons through the inner mitochondrial membrane (review in Klingenberg et al., 1999, Biochim. Biophys. Acta, 1415(2):271-96). The energy stored in the proton gradient is thereby released as heat and not used for ATP synthesis.

- 2 -

When the energy intake of an animal exceeds expenditure surplus energy can be stored as fat in adipose tissue. The generation of a proton leak across the inner mitochondrial membrane by the activation of uncoupling proteins would reduce caloric efficiency and thus avoid the accumulation of excess body fat (obesity) that is detrimental to the animals health. In human, however, brown adipose tissue is almost absent in adults. Therefore, UCP1 was not considered to be a major factor in the formation or prevention of human obesity. Recently, the discovery of further proteins of similar sequence (UCP2-UCP5) that are widely expressed in human tissues (e.g. white adipose tissue, muscle) made this members of the UCP family to important targets for pharmaceutical research (reviewed in Adams 2000, Nutr., 130(4):711-4). Interestingly, and as reviewed in Ricquier, 2000, Biochem J. 345, 161-179, further homologues have been identified, like, inter alia, the plant UCPs StUCP (from *Solanum tuberosum*) and AtUCP (*Arabidopsis thaliana*). Although the in vivo function of these proteins is still unknown, the possibility to influence UCP activity would be a conceivable therapy for the treatment or prevention of obesity and related diseases.

There are several metabolic diseases of human and animal metabolism, e.g., obesity and severe weight loss, that relate to energy imbalance where caloric intake versus energy expenditure is imbalanced. Obesity is one of the most prevalent metabolic disorders in the world. It is a still poorly understood human disease that becomes more and more relevant for western society. Obesity is defined as an excess of body fat, frequently resulting in a significant impairment of health. Besides severe risks of illness such as diabetes, hypertension and heart disease, individuals suffering from obesity are often isolated socially. Human obesity is strongly influenced by environmental and genetic factors, whereby the environmental influence is often a hurdle for the identification of (human) obesity genes. Obesity is influenced by genetic, metabolic, biochemical, psychological, and behavioral factors. As such, it is a complex disorder

- 3 -

that must be addressed on several fronts to achieve lasting positive clinical outcome. Obese individuals are particularly prone to ailments including: diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, osteoarthritis and gallstones.

5

Hyperlipidemia and elevation of free fatty acids correlate clearly with the 'Metabolic Syndrome'. The concept of metabolic syndrome (syndrome x, insulin-resistance syndrome, deadly quartet) was first described 1966 by Camus and reintroduced 1988 by Reaven (Camus JP, 1966, Rev Rhum Mal Osteoartic 33(1):10-14; Reaven et al. 1988, Diabetes, 37(12):1595-1607). Today "metabolic syndrome" is commonly defined as clustering of cardiovascular risk factors like hypertension, abdominal obesity, high blood levels of triglycerides and fasting glucose as well as low blood levels of HDL cholesterol. Insulin resistance greatly increases the risk of developing the metabolic syndrome (Reaven, 2002, Circulation 106(3): 286-8 reviewed). The metabolic syndrome often precedes the development of type II diabetes and cardiovascular disease (McCook, 2002, JAMA 288:2709-2716).

20 Obesity is not to be considered as a single disorder but a heterogeneous group of conditions with (potential) multiple causes. Obesity is also characterized by elevated fasting plasma insulin and an exaggerated insulin response to oral glucose intake (Koltermann, J. Clin. Invest 65, 1980, 1272-1284) and a clear involvement of obesity in type 2 diabetes mellitus can be confirmed (Kopelman, Nature 404, 2000, 635-643).

Even if several candidate genes have been described which are supposed to influence the homeostatic system(s) that regulate body mass/weight, like leptin, VCPI, VCPL, or the peroxisome proliferator-activated receptor-gamma co-activator, the distinct molecular mechanisms and/or molecules influencing obesity or body weight/body mass regulations are not known.

30

- 4 -

Mitochondria have a very specialized function in energy conversion and said function is reflected in their morphological structure, namely the distinct internal membrane. This internal membrane does not only provide the framework for electron-transport processes but also creates a large internal compartment in each organelle in which highly specialized enzymes are confined. Therefore, there is a strong relationship between mitochondrial energy metabolism and the biochemical/biophysical properties of these organelles.

The technical problem underlying the invention was to provide for means and methods for modulating the biological/biochemical activities of mitochondria and, thereby, modulating metabolic conditions in eukaryotic cells which influence energy expenditure, body temperature, thermogenesis, cellular metabolism to an excessive or deficient supply of substrate(s) in order to regulate the ATP level, the  $\text{NAD}^+/\text{NADH}$  ratio, and/or superoxide production. The solution to this technical problem is achieved by providing the embodiments characterized in the claims.

As shown in the appended examples, this invention discloses genes that can suppress the eye defect induced by the activity of dUCPy. These genes are coding for cornichon (GadFly Accession Number CG5855), neuralized (GadFly Accession Number CG11988), dco (GadFly Accession Number CG2048), kraken (GadFly Accession Number CG3943), escargot (GadFly Accession Number CG3758), GadFly Accession Number CG11940, dappled (GadFly Accession Number CG1624), GadFly Accession Number CG11753, GadFly Accession Number CG7262, GadFly Accession Number CG4291. In addition, as shown in the appended examples, this invention discloses genes that can enhance the eye defect induced by the activity of dUCPy. These genes are coding for GadFly Accession Number CG8479, Imp (GadFly Accession Number CG1691), GadFly Accession Number CG8311, Gdh (GadFly Accession Number CG5320), Sir2 (GadFly Accession Number CG5216), msl-2 (GadFly

- 5 -

Accession Number CG3241). It is envisaged that mutations in one or several of these genes affect the activity of uncoupling proteins (UCPs) thereby leading to an altered mitochondrial activity. The present invention provides for specific genes involved in the regulation of diseases and disorders related to body-weight regulation and thermogenesis, for example, but not limited to, metabolic diseases such as obesity, as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis and gallstones and disorders related to ROS defence, such as diabetes mellitus and neurodegenerative disorders.

The term 'GenBank Accession number' relates to NCBI GenBank database entries (Benson et al, Nucleic Acids Res. 28, 2000, 15-18).

The *Drosophila* gene with GadFly Accession Number CG8479 encodes for a protein which is most homologous to human OPA1, optic atrophy 1 (KIAA0567) protein (SEQ ID NO: 4; predicted coding nucleotide sequence; SEQ ID NO: 5; protein; GenBank Accession Number XP\_039926.2) and to mouse large GTP binding protein (Accession Number BAB59000.1). Dominant optic atrophy is the commonest form of inherited optic neuropathy.

The *Drosophila* gene with GadFly Accession Number CG5855 encodes for protein which is most homologous to human cornichon-like protein (SEQ ID NO: 6; predicted coding nucleotide sequence; SEQ ID NO:7; protein; GenBank Accession Number NP\_005767) and to mouse gene Accession Number sp035372. Cornichon, a transmembrane protein, has a crucial but so far undefined role in epidermal growth factor (EGF) signaling during *Drosophila* embryogenesis. Human cornichon which is expressed in a variety of human tissues functions in similar signaling establishing vectorial re-localization and concentration of signaling events in T-cell activation (Utku, 1999, Biochim Biophys Acta;1449(3):203-10).

- 6 -

The *Drosophila* gene with GadFly Accession Number CG1691 encodes for a protein which most homologous to human IGF-II mRNA-binding protein 3 (SEQ ID NO: 8; predicted coding nucleotide sequence; SEQ ID NO: 9; protein; GenBank Accession Number NP\_006538.1) and to mouse gene  
5 with GenBank Accession Number NP\_034081.1. Human IGF (insulin growth factor)-II mRNA binding proteins are major fetal growth factors implicated in rRNA localization and translational control vertebrate development.

10 The *Drosophila* gene neuralized (neur) with GadFly Accession Number CG11988 encodes for a protein which is homologous to human neuralized-like protein (GenBank Accession Number NP\_004201.1 for the protein (SEQ ID NO:11), NM\_004210 for the cDNA (SEQ ID NO:10)). The *Drosophila* neurogenic gene neuralized is expressed in precursors of larval  
15 and adult neurons, embryonic mesoderm and specific follicle cells in the ovary (Boulianne G.L. et al., 1991, EMBO J 10(10):2975-2983). The protein neuralized is necessary for Notch activation. In *Drosophila*, neuralized encodes a peripheral membrane protein involved in delta signaling and endocytosis (Pavlopoulos E. et al., 2001, Dev Cell  
20 1(6):807-816). *Xenopus* neuralized (Xneur) is a ubiquitin ligase that interacts with Xdelta 1 and regulates Notch signaling (Deblandre G.A. et al, 2001, Dev Cell 1(6):795-806). XNeur plays a conserved role in Notch activation by regulating the cell surface levels of the Delta ligands via ubiquitination. h-neu (human neuralized) encodes a protein with strong  
25 homology to the *Drosophila* neuralized (D-neu) protein. The h-neu gene plays a role in determination of cell fate in the human central nervous system and may act as a tumor suppressor whose inactivation could be associated with malignant progression of astrocytic tumors (Nakamura H. et al., 1998, Oncogene 16(8):1009-1019).

30 The *Drosophila* gene with GadFly Accession Number CG8311 encodes for a protein, which is most homologous to human KIAA1094 protein (SEQ ID

- 7 -

NO: 13; GenBank Accession Number NP\_055723.1 for the protein (SEQ ID NO: 12, NM\_014908 for the cDNA), which is a transmembrane protein located in the plasma membrane (PsortII prediction, 74%). No functional data have been published for this protein.

5

The casein kinase I (CKI) family of protein kinases is a group of highly related, ubiquitously expressed serine/threonine kinases found in all eukaryotic organisms from protozoa to man. (Vielhaber and Virshup, 2001, IUBMB Life 51(2):73-78) Recent advances in diverse fields, including  
10 developmental biology and chronobiology, have elucidated roles for CKI in regulating critical processes such as Wnt signaling, circadian rhythm, nuclear import, and Alzheimer's disease progression. Casein kinase I is a serine/threonine-specific protein kinase that constitutes most of the kinase activity in eukaryotic cells, where it is mainly localized in the nucleus,  
15 cytoplasm, and several membranes. The monomeric enzyme phosphorylates hierarchically a variety of substrates without the involvement of the second messenger in signal transduction.

Drosophila double-time (dbt) gene, which encodes a protein similar to  
20 vertebrate epsilon and delta isoforms of casein kinase I, is essential for circadian rhythmicity because it regulates the phosphorylation and stability of period (per) protein (Bao et al. 2001, J Neurosci 21(18):7117-26). Lee et al have provided in vivo evidence that, in addition to casein kinase I epsilon, casein kinase I delta is a second clock relevant kinase (2001, Cell  
25 107(7):855-67). The human casein kinase I delta nucleotide sequence is shown in SEQ ID NO: 14; the amino acid sequence is shown in SEQ ID NO: 15. The human casein kinase I epsilon nucleotide sequence is shown in SEQ ID NO: 16; the amino acid sequence is shown in SEQ ID NO: 17.

30 The canonical Wnt-signaling pathway is critical for many aspects of development, and mutations in components of the Wnt pathway are carcinogenic. Sufficiency tests identified casein kinase I epsilon

- 8 -

(CKIepsilon) as a positive component of the canonical Wnt/beta-catenin pathway, and necessity tests showed that CKIepsilon is required in vertebrates to transduce Wnt signals (McKay et al., 2001, Dev Biol 235(2):388-396). In addition to CKIepsilon, the CKI family includes several  
5 other isoforms (alpha, beta, gamma, and delta) and their role in Wnt sufficiency tests had not yet been clarified. All wild-type CKI isoforms activate Wnt signaling.

Casein kinase I delta (CKIdelta) and casein kinase I epsilon (CKIepsilon)  
10 have been implicated in the response to DNA damage, but the understanding of how these kinases are regulated remains incomplete. In vitro, these kinases rapidly autophosphorylate, predominantly on their carboxyl-terminal extensions, and this autophosphorylation markedly inhibits kinase activity (Cegielska et al., 1998, J. Biol. Chem.  
15 273:1357-1364).

Glutamate dehydrogenase (GDH) is an enzyme catalyzing the oxidative deamination of glutamate to alpha-ketoglutarate using NAD or NADP as cofactors. In mammalian brain, GDH is located predominantly in astrocytes,  
20 where it is involved in the metabolism of neurotransmitter glutamate (see, for example, Plaitakis and Zaganas, 2001, J Neurosci Res 1;66(5):899-908). In human, GDH exists in two isoforms, encoded by the GLUD1 (referred to as housekeeping) and GLUD2 (referred to as nerve tissue-specific) genes which differ in their catalytic and allosteric  
25 properties. The housekeeping GDH is regulated primarily by GTP, the nerve tissue GDH activity depends largely on available ADP or L-leucine levels. Interestingly, the uncoupling protein - 1 (referred to as UCP-1) is also regulated by these nucleotides but adversely to the nerve tissue-specific GDH; ADP inactivates and GTP activates UCP-1. The human glutamate  
30 dehydrogenase I nucleotide sequence is shown in SEQ ID NO: 18; the amino acid sequence is shown in SEQ ID NO: 19. The human glutamate

- 9 -

dehydrogenase II nucleotide sequence is shown in SEQ ID NO: 20; the amino acid sequence is shown in SEQ ID NO: 21.

Glutamate is the precursor of the inhibitory neurotransmitter GABA.

5 Disruptions of glutamate metabolism have been implicated in clinical disorders, such as, for example congenital hyperinsulinism and pyridoxine-dependent seizures. The hyperinsulinism/hyperammonemia syndrome is a form of congenital hyperinsulinism in which children have hypoglycemia together with elevations of plasma ammonium levels. The  
10 disorder is caused by dominant mutations of the mitochondrial GDH, that impair sensitivity to the allosteric inhibitor GTP (see, for example, MacMullen et al., 2001, J Clin Endocrinol Metab 86(4):1782-7). Congenital hyperinsulinism is thus implicating a role of glutamate oxidation by GDH in beta-cell insulin secretion and in hepatic and CNS ammonia detoxification  
15 (see, for example, Kelly and Stanley, 2001, Ment Retard Dev Disabil Res Rev 2001;7(4):287-95).

Dietary-induced obesity in rats showed a stable, higher body weight than controls, and key enzymes of alpha-amino nitrogen metabolism, including  
20 glutamine synthetase and GDH, showed reduced activities in brown adipose tissue of obese rats (see, for example, Serra et al., 1994, Biochem Mol Biol Int 32(6):1173-1188). These adaptations in amino acid metabolism were dependent on the obese status of the rats.

25 The Drosophila gene kraken with GadFly Accession Number CG3943 encodes for a protein which is most homologous to protein encoded by a novel human gene mapping to chromosome 22 (SEQ ID NO:23; GenBank Accession Number CAC16804.1 for the protein, SEQ ID NO: 22; AL450314 for the cDNA). No functional data are available for this protein.

30 The Drosophila gene with GadFly Accession Number CG5216 encodes for Sir2 (also referred to as sirtuin) protein. Sir2 protein is most homologous to

- 10 -

human Sirtuin 1 protein (SEQ ID NO: 24; predicted coding nucleotide sequence; SEQ ID NO:25; protein; GenBank Accession Number NP\_036370) and to mouse Sirtuin 1 protein (GenBank Accession Number NP\_062786.1). Sirtuins (silent mating type information regulation) are a large family of NAD-dependent deacetylase enzymes. These proteins are conserved from prokaryotes to eukaryotes, but most remain uncharacterized, including all seven human sirtuins (Grotzinger et al., 2001, J Biol Chem 276(42):38837-43).

The *Drosophila* *esg* gene with GadFly Accession Number CG3758 encodes for escargot (also referred to as Esgarot) protein, a specific RNA polymerase II transcription factor which is a component of the nucleus. *Drosophila* *esg* is a key regulator of cell adhesion and motility in tracheal morphogenesis. *Esg* is most homologous to human hypothetical protein, similar to gonadotropin protein (SEQ ID NO: 26; predicted coding nucleotide sequence; SEQ ID NO:27; protein; GenBank Accession Number XP\_030528) and to mouse gene with the Accession Number NP\_035545. No functional data are available for the mammalian proteins.

The *Drosophila* gene with GadFly Accession Number CG3241 encodes for *msl-2* (male specific lethal 2) protein. *Msl-2* protein is most homologous to human hypothetical KIAA1585 protein (SEQ ID NO: 28; predicted coding nucleotide sequence; SEQ ID NO:29; protein; GenBank Accession Number AB046805) and to mouse protein with GenBank Accession Number BF471233. The *Drosophila* male-specific lethal (MSL) genes regulate transcription from the male X chromosome in a dosage compensation pathway that equalizes X-linked gene expression in males and females. *Drosophila* *Msl-2* is part of a protein complex that regulates gene activities by altering the chromatin structure (Kageyama et al., 2001, EMBO J 20(9):2236-45). Zhou et al. described that the *Drosophila* male-specific lethal 2 (*msl-2*) gene is involved in dosage compensation (1995, EMBO J 14(12):2884-95). The encoded protein (MSL-2) has a RING finger (C3HC4

- 11 -

zinc finger) and a metallothionein-like domain and undergoes sex-specific regulation. The protein Sex-lethal (SXL) controls dosage compensation in *Drosophila* by inhibiting splicing and subsequently translation of male-specific-lethal-2 (*msl-2*) transcripts (Forch et al., 2001, RNA 7(9):1185-91).

The *Drosophila* gene with GadFly Accession Number CG11940 encodes for alsin protein. Alsin protein is most homologous to human Alsin aslrcr9 protein (SEQ ID NO: 30; predicted coding nucleotide sequence; SEQ ID NO:31; protein; GenBank Accession Number XP\_028059.1) and to mouse Alsin protein (GenBank Accession Number AAH03991). Alsin, a protein with three guanine-nucleotide (GTP) exchange factor domains, has been identified to be responsible for amyotrophic lateral sclerosis which is a neurodegenerative condition that affects large motor neurons of the central nervous system.

The *Drosophila* gene dappled (*dpld*) with GadFly Accession Number CG1624 encodes for a protein which is most homologous to human protein (SEQ ID NO:33; GenBank Accession Number XP\_067369.1 for the protein, SEQ ID NO: 32; XM\_067369 for the cDNA), similar to C12C8.3b.p. No functional data are available for the human protein. C12C8.3b.p is a *Caenorhabditis elegans* protein with GenBank Accession Number NP\_492488.

The *Drosophila* gene with GadFly Accession Number CG11753 encodes for a protein which is most homologous to human protein (SEQ ID NO:35; GenBank Accession Number XP\_029849.1 for the protein, SEQ ID NO: 34; XM\_029849 for the cDNA), encoded by a gene similar to mouse RIKEN cDNA 2610042O14 gene (GenBank Accession Number NM\_025575). No functional data are available for these proteins.

- 12 -

The *Drosophila* gene with GadFly Accession Number CG7262 encodes for a protein which is most homologous to human KIAA0095 protein (SEQ ID NO:37; GenBank Accession Number NP\_055484.1 for the protein; SEQ ID NO: 36; NM\_014669 for the cDNA (Nagase et al., 1995, DNA Res. 2 (1):37-43); GenBank Accession Number AX306779, Sequence 12 from Patent WO0018961). No functional data are available for this protein. The KIAA0095 gene is related to *S. cerevisia* NIC96 gene (GenBank Accession Number P34077) which is part of the nucleoporin complex and is required for protein transport in the nucleus. The KIAA0095 protein also shows homologies to *Xenopus* An4a protein (GenBank Accession Number AAB49669) and Zebrafish hi4 "dead eye" protein (GenBank Accession Number AAB61137).

The *Drosophila* gene with GadFly Accession Number CG4291 encodes for a protein which is most homologous to human WW domain binding protein 4 (formin binding protein 21 (FBP21); SEQ ID NO: 38; predicted coding nucleotide sequence; SEQ ID NO:39; protein; GenBank Accession Number XP\_049375) and to mouse WW domain binding protein 4 (formin binding protein 21) gene with the Accession Number NP\_061235. The WW domain is a protein module with two highly conserved tryptophans that binds proline-rich peptide motifs in vitro. The *Drosophila* gene CG4291 encodes a small nuclear ribonucleoprotein involved in mRNA splicing which is a component of the snRNP U2e. Human FBP21 is present in highly purified spliceosomal complex A, is associated with U2 snRNPs, and colocalizes with splicing factors in nuclear speckle domains. FBP21 may play a role in cross-intron bridging of snRNPs in the mammalian A complex.

So far, it has not been described that Optic atrophy 1 protein (OPA1), cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta and epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein,

- 13 -

formin-binding protein 21, or a homologous protein is involved in the regulation of body-weight and thermogenesis, for example, but not limited to, metabolic diseases such as obesity, as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, and gallstones, and disorders related to ROS defence, such as diabetes mellitus and neuro-degenerative disorders, and thus, no functions in metabolic diseases and other diseases as listed above have been discussed in the prior art.

In this invention we demonstrate that the correct gene dose of the *Drosophila melanogaster* homologues of SEQ ID NO:4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, or 38 is essential for maintenance of energy homeostasis and for the activity of mitochondrial uncoupling protein. A genetic screen was designed to identify factors that modulate activity of uncoupling protein. We discovered that mutation of these genes caused a reduction of the activity of uncoupling protein, thereby leading to an altered mitochondrial activity. Thus, the invention is also based on the finding that homologues of the above *Drosophila* genes, particularly the human homologues as described in SEQ ID NO:4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, and 38 are contributing to membrane stability and/or function of organelles, preferably mitochondria and thus represent targets for diagnostic and/or therapeutic applications in medicine, particularly in human medicine.

The function of the proteins of the invention in metabolic disorders is further validated by data obtained from additional screens. For example, the content of triglycerides and glycogen of a pool of flies with the same genotype was analyzed using a triglyceride and a glycogen assay. Additionally expression profiling studies (see Examples for more detail) confirm the particular relevance of the proteins of the invention as

- 14 -

regulators of energy metabolism in mammals. These findings suggest the presence of similar activities of these described homologous proteins in humans that provides insight into diagnosis, treatment, and prognosis of metabolic disorders.

5

Polynucleotides encoding proteins as shown in SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, and 39 are suitable to investigate, to treat, to prevent or to diagnose diseases and disorders related to body-weight regulation and thermogenesis, for example, but not limited to, metabolic diseases such as obesity, as well as related disorders as described above. Molecules related to SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, or 39 provide new compositions useful in diagnosis, treatment, and prognosis of diseases and disorders related to body-weight regulation and thermogenesis as described above.

10

15

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention, which will be limited only by the appended claims. Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing the cell lines, vectors, and methodologies, which are reported in the publications which might be used in connection with the invention. Nothing herein is to be

20

25

30

- 15 -

construed as an admission that the invention is not entitled to antedate such disclosure.

The present invention discloses that the proteins as shown in SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, and 39 and homologous proteins are directly or indirectly involved in membrane stability and/or function of organelles, in particular mitochondria, and polynucleotides, which identify and encode the proteins are disclosed in this invention. The invention also relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides of the invention. The invention also relates to the use of these sequences and effectors thereof, e.g. antibodies, aptamers or other receptors recognizing the nucleic acid molecules or polypeptides encoded thereby in the diagnosis, study, prevention, and treatment of diseases and disorders related to body-weight regulation and thermogenesis, for example, but not limited to, metabolic diseases such as obesity, as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis and gallstones and disorders related to ROS defence, such as diabetes mellitus and neurodegenerative disorders.

The invention relates to a pharmaceutical composition comprising a nucleic acid molecule of the Optic atrophy 1 protein (OPA1), cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta and epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, or formin-binding protein 21 gene family or a polypeptide encoded thereby or a fragment or a variant of said nucleic acid molecule or said polypeptide or an antibody, an aptamer or another receptor recognizing said nucleic acid molecule or a polypeptide encoded thereby together with pharmaceutically acceptable carriers, diluents and/or adjuvants.

- 16 -

Proteins as shown in SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, and 39 and homologous proteins and nucleic acid molecules coding therefore are obtainable from insect or vertebrate species, e.g. mammals or birds. Particularly preferred are human nucleic acid molecules as shown in SEQ ID NO:4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, or 38 (Optic atrophy 1 protein (OPA1), cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta and epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, formin-binding protein 21, and homologous proteins), i.e. nucleic acids encoding a the proteins of SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, or 39.

The invention particularly relates to a nucleic acid molecule encoding a polypeptide contributing to regulating the energy homeostasis, and/or contributing to membrane stability and/or function of organelles, wherein said nucleic acid molecule comprises

- (a) a nucleotide sequence as shown in SEQ ID NO:4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, or 38 and/or a nucleotide sequence complementary thereto,
- (b) a nucleotide sequence which hybridizes at 50°C in a solution containing 1 x SSC to a nucleic acid molecule encoding an amino acid sequence as shown in SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, or 39 and/or a nucleic acid molecule complementary thereto,
- (c) a sequence corresponding to the sequences of (a) or (b) within the degeneration of the genetic code,
- (d) a sequence which encodes a polypeptide which is at least 85%, preferably at least 90%, more preferably at least 95%, more preferably at least 98% and up to 99,6% identical to the amino acid

- 17 -

sequences as shown in SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, or 39;

- (e) a sequence which differs from the nucleic acid molecule of (a) to (d) by mutation and wherein said mutation causes an alteration, deletion, duplication or premature stop in the encoded polypeptide or
- (f) a partial sequence of any of the nucleotide sequences of (a) to (e) having a length of at least 15 bases, preferably at least 20 bases, more preferably at least 25 bases and most preferably at least 50 bases.

The present invention discloses that the proteins as shown in SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, and 39 and homologous proteins are directly or indirectly involved in membrane stability and/or function of organelles, in particular mitochondria, and polynucleotides, which identify and encode the proteins disclosed in this invention. The invention describes the use of these compositions for the diagnosis, study, prevention, or treatment of diseases and disorders related to body-weight regulation and thermogenesis as described above.

The ability to manipulate and screen the genomes of model organisms such as the fly *Drosophila melanogaster* provides a powerful tool to analyze biological and biochemical processes that have direct relevance to more complex vertebrate organisms due to significant evolutionary conservation of genes, cellular processes, and pathways (see, for example, Adams M. D. et al., (2000) *Science* 287: 2185-2195). Identification of novel gene functions in model organisms can directly contribute to the elucidation of correlative pathways in mammals (humans) and of methods of modulating them. A correlation between a pathology model and the modified expression of a fly gene can identify the association of the human ortholog with the particular human disease.

- 18 -

Accordingly, the present invention relates to genes with novel functions in body-weight regulation, energy homeostasis, metabolism, and obesity. To find genes with novel functions in energy homeostasis, metabolism, and obesity, a functional genetic screen was performed with the model organism *Drosophila melanogaster* (Meigen). One resource for screening was a proprietary *Drosophila melanogaster* stock collection of EP-lines. The P-vector of this collection has Gal4-UAS-binding sites fused to a basal promoter that can transcribe adjacent genomic *Drosophila* sequences upon binding of Gal4 to UAS-sites. This enables the EP-line collection for overexpression of endogenous flanking gene sequences. In addition, without activation of the UAS-sites, integration of the EP-element into the gene is likely to cause a reduction of gene activity, and allows determining its function by evaluating the loss-of-function phenotype.

It is preferred that the nucleic acid molecule encodes a polypeptide contributing to membrane stability and/or function of organelles and represents a protein of *Drosophila* which has been found to be able to modify UCPs, see also appended examples. As demonstrated in the appended examples, the here described polypeptide (and encoding nucleic acid molecule) was able to modify, e.g. suppress or enhance a specific eye phenotype in *Drosophila* which was due to the overexpression of the *Drosophila melanogaster* gene dUCPy. The overexpression of dUCPy (with homology to human UCPs) in the compound eye of *Drosophila* led to a clearly visible eye defect which can be used as a 'read-out' for a genetical 'modifier Screen'.

In said "modifier screen" thousands of different genes are mutagenized to modify their expression in the eye. Should one of the mutagenized genes interact with dUCPy and modify its activity an enhancement or suppression of the eye defect will occur. Since such flies are easily to discern they can be selected to isolate the interacting gene. As shown in the appended examples, genes were deduced that can enhance or suppress the eye

- 19 -

defect induced by the activity of dUCPy. The identified genes have high homologies to the human proteins shown in SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, and 39, as described above. It is envisaged that mutations in the herein described proteins (and  
5 corresponding genes) lead to phenotypic and/or physiological changes which may comprise a modified and altered mitochondrial activity. This, in turn, may lead to, inter alia, an altered energy metabolism, altered thermogenesis and/or altered energy homeostasis.

10 As shown in the appended examples, new genes were found that can enhance or suppress the eye defect induced by the activity of dUCPy. The genes suppressing the eye defect are cornichon (GadFly Accession Number CG5855), neuralized (GadFly Accession Number CG11988), dco (GadFly Accession Number CG2048), kraken (GadFly Accession Number CG3943),  
15 escargot (GadFly Accession Number CG3758), GadFly Accession Number CG11940, dappled (GadFly Accession Number CG1624), GadFly Accession Number CG11753, GadFly Accession Number CG7262, and GadFly Accession Number CG4291; and the genes enhancing the eye defect induced by UCP activity are GadFly Accession Number CG8479,  
20 Imp (GadFly Accession Number CG1691), GadFly Accession Number CG8311, Gdh (GadFly Accession Number CG5320), Sir2 (GadFly Accession Number CG5216), and msl-2 (GadFly Accession Number CG3241). The invention also encompasses polynucleotides that encode the proteins as shown in SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25,  
25 27, 29, 31, 33, 35, 37, and 39 or homologous proteins. Accordingly, any nucleic acid sequence, which encodes the amino acid sequences of SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, or 39 can be used to generate recombinant molecules that express the corresponding mRNA and protein.

30 In an additional screen using *Drosophila* mutants, the content of triglycerides and glycogen was analyzed after feeding for six days using a

- 20 -

triglyceride and a glycogen assay. Male flies homozygous for the integration of vectors for *Drosophila* lines HD-EP20292, HD-35207, HD-EP20506, HD-EP20817, HD-EP26792, HD-EP25097, and HD-EP10934 were analyzed in assays measuring the triglyceride and glycogen contents of these flies, illustrated in more detail in the EXAMPLES section. The results of the triglyceride and glycogen content analysis are shown in FIGURES 6, 16, 20, 22, and 23D.

Expression profiling studies (see Examples for more detail) confirm the particular relevance of the proteins of the invention as regulators of energy metabolism in mammals. OPA1 is expressed in different mammalian tissues, showing 2 to 3 fold higher levels of expression in BAT, hypothalamus, brain, muscle and heart when compared to other tissues (see FIGURE 4A). BAT, brain, muscle and heart represent tissues with the major catabolic activity in the body. The high expression levels of OPA-1 in these tissues indicate, that OPA-1 is involved in the metabolism of tissues relevant for the metabolic syndrome. Neuralized-like is highly expressed in muscle, hypothalamus, brain and testis (see FIGURE 9). The high expression levels in muscle when compared to other tissues is indicative for a role in the metabolism in one of the major catabolic tissues of the body. The CG8311 homologous protein shows highest expression levels in brown adipose tissue compared to several other mouse tissues and organs (see FIGURE 11).

In a particular embodiment, the invention encompasses a polynucleotide comprising the nucleic acid sequence of SEQ ID NO:4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, or 38. It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding the proteins of SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, or 39, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention

- 21 -

contemplates each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the nucleotide sequences as shown in SEQ ID NO:4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, and 38, and all such variations are to be considered as being specifically disclosed. Although nucleotide sequences which encode the proteins of SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, or 39 and variants thereof are preferably capable of hybridising to the nucleotide sequences of the naturally occurring nucleic acids of SEQ ID NO:4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, or 38 under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding Optic atrophy 1 protein (OPA1), cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta and epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, formin-binding protein 21, or homologous proteins or their derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilised by the host. Other reasons for substantially altering the nucleotide sequence without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequences. The invention also encompasses production of DNA sequences, or portions thereof, which encode the proteins of SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, or 39 and derivatives, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art at the time of the filing of this

- 22 -

application. Moreover, synthetic chemistry may be used to introduce mutations into the sequence in any portion thereof.

Also encompassed by the invention are polynucleotide sequences that are  
5 capable of hybridising to the claimed nucleotide sequences, under various  
conditions of stringency. Hybridisation conditions are based on the melting  
temperature ( $T_m$ ) of the nucleic acid binding complex or probe, as taught  
in Wahl, G. M. and S. L. Berger (1987: Methods Enzymol. 152:399-407)  
and Kimmel, A. R. (1987; Methods Enzymol. 152:507-511), and may be  
10 used at a defined stringency. Preferably, hybridization under stringent  
conditions means that after washing for 1 h with 1 x SSC and 0.1% SDS  
at 50°C, preferably at 55°C, more preferably at 62°C and most preferably  
at 68°C, particularly for 1 h in 0.2 x SSC and 0.1% SDS at 50°C,  
preferably at 55°C, more preferably at 62°C and most preferably at 68°C,  
15 a positive hybridization signal is observed. Altered nucleic acid sequences  
encoding the proteins of SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23,  
25, 27, 29, 31, 33, 35, 37, or 39 which are encompassed by the  
invention include deletions, insertions, or substitutions of different  
nucleotides resulting in a polynucleotide that encodes the same or a  
20 functionally equivalent protein.

The encoded proteins may also contain deletions, insertions, or  
substitutions of amino acid residues, which produce a silent change and  
result in a functionally equivalent Optic atrophy 1 protein (OPA1),  
25 cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094  
protein, casein kinase (delta and epsilon), glutamate dehydrogenase,  
kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940  
homolog, dappled homolog, CG11753 homolog, KIAA0095 protein,  
formin-binding protein 21, or homologous protein. Deliberate amino acid  
30 substitutions may be made on the basis of similarity in polarity, charge,  
solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of  
the residues as long as the biological activity is at least partially retained.

- 23 -

For example, negatively charged amino acids may include aspartic acid and glutamic acid; positively charged amino acids may include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine  
5 and alanine; asparagine and glutamine; serine and threonine; phenylalanine and tyrosine. Furthermore, the invention relates to peptide fragments of the proteins or derivatives thereof such as cyclic peptides, retro-inverso peptides or peptide mimetics having a length of at least 4, preferably at least 6 and up to 50 amino acids.

10 Also included within the scope of the present invention are alleles of the genes encoding the proteins of SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, or 39. As used herein, an "allele" or "allelic sequence" is an alternative form of the gene, which may result from  
15 at least one mutation in the nucleic acid sequence. Alleles may result in altered mRNAs or polypeptides whose structures or function may or may not be altered. Any given gene may have none, one, or many allelic forms. Common mutational changes, which give rise to alleles, are generally ascribed to natural deletions, additions, or substitutions of nucleotides.  
20 Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

The nucleic acid sequences encoding SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, or 39 may be extended utilising a  
25 partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one method which may be employed, "restriction-site" PCR, uses universal primers to retrieve unknown sequence adjacent to a known locus (Sarkar, G. (1993) PCR Methods  
30 Applic. 2:318-322). Inverse PCR may also be used to amplify or extend sequences using divergent primers based on a known region (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186).

- 24 -

Another method which may be used is capture PCR which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (PCR Methods Applic. 1:111-119). Another method which may be used to retrieve  
5 unknown sequences is that of Parker, J. D. et al. (1991; Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries to walk in genomic DNA (Clontech, Palo Alto, Calif.). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

10 In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode the proteins of SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, or 39, or fusion proteins or functional equivalents thereof, may be used in recombinant  
15 DNA molecules for the expression of the proteins in appropriate host cells.

In order to express a biologically active protein, the nucleotide sequences encoding the proteins or functional equivalents, may be inserted into appropriate expression vectors, i.e., a vector which contains the necessary  
20 elements for the transcription and translation of the inserted coding sequence. Methods, which are well known to those skilled in the art, may be used to construct expression vectors containing sequences encoding the proteins and the appropriate transcriptional and translational control elements. Regulatory elements include for example a promoter, an initiation  
25 codon, a stop codon, a mRNA stability regulatory element, and a polyadenylation signal. Expression of a polynucleotide can be assured by (i) constitutive promoters such as the Cytomegalovirus (CMV) promoter/enhancer region, (ii) tissue specific promoters such as the insulin promoter (see, Soria et al., 2000, Diabetes 49:157), SOX2 gene promotor  
30 (see Li et al., (1998) Curr. Biol. 8:971-4), Msi-1 promotor (see Sakakibara et al., (1997) J. Neuroscience 17:8300-8312), alpha-cardia myosin heavy chain promotor or human atrial natriuretic factor promotor (Klug et al.,

- 25 -

(1996) J. clin. Invest 98:216-24; Wu et al., (1989) J. Biol. Chem. 264:6472-79) or (iii) inducible promoters such as the tetracycline inducible system. Expression vectors can also contain a selection agent or marker gene that confers antibiotic resistance such as the neomycin, hygromycin  
5 or puromycin resistance genes. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. Such techniques are described in Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y. and Ausubel, F.M. et al. (1989) Current Protocols in Molecular Biology,  
10 John Wiley & Sons, New York, N.Y.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding the proteins of the invention and homologous proteins may be ligated to a heterologous sequence to encode  
15 a fusion protein.

In order to express biologically active proteins, the nucleotide sequences coding therefor or for functional equivalents, may be inserted into appropriate expression vectors, i.e. a vector, which contains the necessary  
20 elements for the transcription and translation of the inserted coding sequence. Methods, which are well known to those skilled in the art, may be used to construct expression vectors containing sequences encoding the proteins and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques,  
25 synthetic techniques, and in vivo genetic recombination. Such techniques are described in Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y., and Ausubel, F. M. et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y.

30

A variety of expression vector/host systems may be utilised to contain and express a sequence encoding the proteins of SEQ ID NO:5, 7, 9, 11, 13,

- 26 -

15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, or 39 or fusion proteins. These include, but are not limited to, micro-organisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors;  
5 insect cell systems infected with virus expression vectors (e.g. baculovirus, adenovirus, adeno-associated virus, lentivirus, retrovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g. Ti or PBR322 plasmids); or animal cell systems.

10 The presence of polynucleotide sequences encoding the protein can be detected by DNA-DNA or DNA-RNA hybridisation and/or amplification using probes or portions or fragments of polynucleotides encoding the protein. Nucleic acid amplification based assays involve the use of  
15 oligonucleotides or oligomers based on the sequences encoding the protein to detect transformants containing DNA or RNA encoding the protein. As used herein "oligonucleotides" or "oligomers" refer to a nucleic acid sequence of at least about 10 nucleotides and as many as about 60 nucleotides, preferably about 15 to 30 nucleotides, and more preferably  
20 about 20-25 nucleotides, which can be used as a probe or amplimer.

The presence of proteins of the invention in a sample can be determined by immunological methods or activity measurement. A variety of protocols for detecting and measuring the expression of the proteins using either  
25 polyclonal or monoclonal antibodies specific for the protein or reagents for determining protein activity are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilising monoclonal antibodies reactive to two  
30 non-interfering epitopes on the protein is preferred, but a competitive binding assay may be employed. These and other assays are described, among other places, in Hampton, R. et al. (1990; Serological Methods, a

- 27 -

Laboratory Manual, APS Press, St Paul, Minn.) and Maddox, D. E. et al. (1983; J. Exp. Med. 158:1211-1216).

5 A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labelled hybridisation or PCR probes for detecting sequences related to polynucleotides encoding the protein include oligo-labelling, nick translation, end-labelling or PCR amplification using a labelled nucleotide, or enzymatic synthesis. These procedures may  
10 be conducted using a variety of commercially available kits (Pharmacia & Upjohn, (Kalamazoo, Mich.); Promega (Madison Wis.); and U.S. Biochemical Corp., (Cleveland, Ohio).

15 Alternatively, the sequences encoding the protein, or any portions thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesise RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labelled nucleotides. These procedures may be conducted using a variety of commercially available kits  
20 (Pharmacia & Upjohn, (Kalamazoo, Mich.); Promega (Madison Wis.); and U.S. Biochemical Corp., (Cleveland, Ohio).

Suitable reporter molecules or labels, which may be used, include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic  
25 agents as well as substrates, co-factors, inhibitors, magnetic particles, and the like.

The nucleic acids encoding the proteins of the invention can be used to generate transgenic animal or site specific gene modifications in cell lines.  
30 Transgenic animals may be made through homologous recombination, where the normal locus of the genes encoding the proteins of the invention is altered. Alternatively, a nucleic acid construct is randomly integrated into

- 28 -

the genome. Vectors for stable integration include plasmids, retroviruses and other animal viruses, YACs, and the like. The modified cells or animal are useful in the study of the function and regulation of the proteins of the invention. For example, a series of small deletions and/or substitutions may  
5 be made in the genes that encode the proteins of the invention to determine the role of particular domains of the protein, functions in pancreatic differentiation, etc.

Specific constructs of interest include anti-sense molecules, which will  
10 block the expression of the proteins of the invention, or expression of dominant negative mutations. A detectable marker, such as for example lac-Z, may be introduced in the locus of the genes of the invention, where upregulation of expression of the genes of the invention will result in an easily detected change in phenotype.

15 One may also provide for expression of the genes of the invention or variants thereof in cells or tissues where it is not normally expressed or at abnormal times of development. In addition, by providing expression of the proteins of the invention in cells in which they are not normally produced,  
20 one can induce changes in cell behavior.

DNA constructs for homologous recombination will comprise at least portions of the genes of the invention with the desired genetic modification, and will include regions of homology to the target locus. DNA  
25 constructs for random integration need not include regions of homology to mediate recombination. Conveniently, markers for positive and/or negative selection are included. Methods for generating cells having targeted gene modifications through homologous recombination are known in the art. For embryonic stem (ES) cells, an ES cell line may be employed, or embryonic  
30 cells may be obtained freshly from a host, e.g. mouse, rat, guinea pig etc. Such cells are grown on an appropriate fibroblast-feeder layer or grown in presence of leukemia inhibiting factor (LIF).

- 29 -

When ES or embryonic cells or somatic pluripotent stem cells have been transformed, they may be used to produce transgenic animals. After transformation, the cells are plated onto a feeder layer in an appropriate medium. Cells containing the construct may be detected by employing a selective medium. After sufficient time for colonies to grow, they are picked and analyzed for the occurrence of homologous recombination or integration of the construct. Those colonies that are positive may then be used for embryo manipulation and blastocyst injection. Blastocysts are obtained from 4 to 6 week old superovulated females. The ES cells are trypsinized, and the modified cells are injected into the blastocoel of the blastocyst. After injection, the blastocysts are returned to each uterine horn of pseudopregnant females. Females are then allowed to go to term and the resulting offspring screened for the construct. By providing for a different phenotype of the blastocyst and the genetically modified cells, chimeric progeny can be readily detected. The chimeric animals are screened for the presence of the modified gene and males and females having the modification are mated to produce homozygous progeny. If the gene alterations cause lethality at some point in development, tissues or organs can be maintained as allogenic or congenic grafts or transplants, or in vitro culture. The transgenic animals may be any non-human mammal, such as laboratory animal, domestic animals, etc. The transgenic animals may be used in functional studies, drug screening, etc.

## Diagnostics and Therapeutics

The data disclosed in this invention show that the nucleic acids and proteins of the invention and effector molecules thereof are useful in diagnostic and therapeutic applications implicated, for example but not limited to, in metabolic disorders like obesity, diabetes, eating disorders, wasting syndromes (cachexia), pancreatic dysfunctions, arteriosclerosis, coronary artery disease (CAD), and other diseases and disorders. Hence,

- 30 -

diagnostic and therapeutic uses for the proteins of the invention, e.g. the proteins of SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, or 39 are, for example but not limited to, the following: (i) protein therapeutic, (ii) small molecule drug target, (iii) antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) diagnostic and/or prognostic marker, (v) gene therapy (gene delivery/gene ablation), (vi) research tools, and (vii) tissue regeneration in vitro and in vivo (regeneration for all these tissues and cell types composing these tissues and cell types derived from these tissues).

The nucleic acids and proteins of the invention are useful in diagnostic and therapeutic applications implicated in various diseases and disorders described below and/or other pathologies and disorders. For example, but not limited to, cDNAs encoding the proteins of the invention and particularly their human homologues may be useful in gene therapy, and the proteins of the invention and particularly their human homologues may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from, for example, but not limited to, in metabolic disorders like obesity, diabetes, eating disorders, wasting syndromes (cachexia), pancreatic dysfunctions, arteriosclerosis, coronary artery disease (CAD), and other diseases and disorders.

The nucleic acid encoding the proteins of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acids or the proteins are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

For example, in one aspect, antibodies which are specific for the protein may be used directly as an antagonist, or indirectly as a targeting or

- 31 -

delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express the protein. The antibodies may be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimerical, single chain, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies, (i.e. those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others, may be immunised by injection with the protein or any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminium hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. Among adjuvants used in human, BCG (Bacille Calmette-Guerin) and *Corynebacterium parvum* are especially preferable. It is preferred that the peptides, fragments, or oligopeptides used to induce antibodies to proteins of SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, or 39 have an amino acid sequence consisting of at least five amino acids, and more preferably at least 10 amino acids.

Monoclonal antibodies to the protein may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Köhler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R. J. et al. *Proc. Natl. Acad. Sci.* 80:2026-2030; Cole, S. P. et al. (1984) *Mol. Cell Biol.* 62:109-120).

- 32 -

In addition, techniques developed for the production of "chimeric antibodies", the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can be used (Morrison, S. L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M. S. et al (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce protein-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D. R. (1991) Proc. Natl. Acad. Sci. 88:11120-3). Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening recombinant immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments, which contain specific binding sites for the protein, may also be generated. For example, such fragments include, but are not limited to, the  $F(ab')_2$  fragments which can be produced by Pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of  $F(ab')_2$  fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W. D. et al. (1989) Science 254:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding and immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between the protein and its specific antibody. A two-site,

- 33 -

monoclonal-based immunoassay utilising monoclonal antibodies reactive to two non-interfering epitopes is preferred, but a competitive binding assay may also be employed (Maddox, supra).

5 In another embodiment of the invention, the polynucleotides encoding the proteins of SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, or 39, or effector nucleic acids such as aptamers, antisense molecules, ribozymes or RNAi molecules may be used for therapeutic purposes. In one aspect, aptamers, i.e. nucleic acid molecules  
10 capable of binding to a target protein and modulating its activity may be obtained by known methods, e.g. by affinity selection of combinatorial nucleic acid libraries.

In a further aspect, antisense to the polynucleotide encoding the protein  
15 may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding the protein. Thus, antisense molecules may be used to modulate the protein activity, or to achieve regulation of gene function. Such technology is now well known in  
20 the art, and sense or antisense oligomers or larger fragments, can be designed from various locations along the coding or control regions of sequences encoding the protein. Expression vectors derived from retroviruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids may be used for delivery of nucleotide sequences to the  
25 targeted organ, tissue or cell population. Methods, which are well known to those skilled in the art, can be used to construct recombinant vectors, which will express antisense molecules complementary to the polynucleotides of the gene encoding the protein. These techniques are described both in Sambrook et al. (supra) and in Ausubel et al. (supra).  
30 Genes encoding the protein and can be turned off by transforming a cell or tissue with expression vectors which express high levels of polynucleotide or fragment thereof which encodes the protein. Such constructs may be

- 34 -

used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more  
5 with a non-replicating vector and even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing antisense molecules, DNA, RNA, or nucleic acid analogues such  
10 as PNA, to the control regions of the gene encoding the protein, i.e. the promoters, enhancers, and introns. Oligonucleotides derived from the transcription initiation site, e.g., between positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it  
15 cause inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J. E. et al. (1994) In; Huber, B. E. and B. I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y.). The  
20 antisense molecules may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyse the specific cleavage of RNA. The mechanism of ribozyme action involves  
25 sequence-specific hybridisation of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Examples, which may be used, include engineered hammerhead motif ribozyme molecules that can be specifically and efficiently catalyse endonucleolytic cleavage of sequences encoding the protein. Specific ribozyme cleavage sites within  
30 any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of

- 35 -

between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing  
5 accessibility to hybridisation with complementary oligonucleotides using ribonuclease protection assays.

Nucleic acid effector molecules such as antisense molecules and ribozymes of the invention may be prepared by any method known in the art for the  
10 synthesis of nucleic acid molecules. These include techniques for chemically synthesising oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding the protein. Such DNA sequences may be incorporated into a variety of  
15 vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesise antisense RNA constitutively or inducibly can be introduced into cell lines, cells, or tissues. RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition  
20 of flanking sequences at the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of non-traditional bases such as inosine, queosine, and  
25 wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognised by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and  
30 equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient.

- 36 -

Delivery by transfection and by liposome injections may be achieved using methods, which are well known in the art. Any of the therapeutic methods described above may be applied to any suitable subject including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may comprise the protein, antibodies to the protein, mimetics, agonists, antagonists; or inhibitors of the protein. The compositions may be administered alone or in combination with at least one other agent, such as stabilising compound, which may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs or hormones. The pharmaceutical compositions utilised in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries, which facilitate processing of the active compounds into preparations which, can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, Pa.).

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective

- 37 -

amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art. For any compounds, the therapeutically effective doses can be estimated initially either in cell culture assays, e.g. of preadipocyte cell lines, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. A therapeutically effective dose refers to that amount of active ingredient, for example the nucleic acids or the proteins or fragments thereof or antibodies against the protein which are effective against a specific condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g. ED<sub>50</sub> (the dose therapeutically effective in 50% of the population) and LD<sub>50</sub> (the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD<sub>50</sub>/ED<sub>50</sub>. Pharmaceutical compositions, which exhibit large therapeutic indices, are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that include the ED<sub>50</sub> with little or no toxicity. The dosage varies within this range depending upon the dosage being employed, the sensitivity of the patient, and the route of administration. The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors, which may be taken into account, include the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and

- 38 -

clearance rate of the particular formulation. Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

In another embodiment, antibodies which specifically bind the protein may be used for the diagnosis of conditions or diseases characterised by or associated with over- or underexpression of the protein, or in assays to monitor patients being treated with the protein, agonists, antagonists or inhibitors. The antibodies useful for diagnostic purposes may be prepared in the same manner as those described above for therapeutics. Diagnostic assays for the protein include methods, which utilise the antibody and a label to detect the protein in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labelled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules, which are known in the art may be used several of which are described above.

A variety of protocols including ELISA, RIA, and FACS for measuring the protein are known in the art and provide a basis for diagnosing altered or abnormal levels of protein expression. Normal or standard values for protein expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to the protein under conditions suitable for complex formation. The amount of standard complex formation may be quantified by various methods, but preferably by photometric means. Quantities of the protein expressed in control and disease, samples e.g. from biopsied tissues are compared with

- 39 -

the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding the protein may be used for diagnostic purposes. The polynucleotides, which may be used, include oligonucleotide sequences, antisense RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of the protein may be correlated with disease. The diagnostic assay may be used to distinguish between absence, presence, and excess gene expression and to monitor regulation of gene expression levels during therapeutic intervention.

In one aspect, hybridisation with probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding the protein and closely related molecules, may be used to identify nucleic acid sequences which encode the protein. The specificity of the probe, whether it is made from a highly specific region, e.g. unique nucleotides in the 5' regulatory region, or a less specific region, e.g. especially in the 3' coding region, and the stringency of the hybridisation or amplification (maximal, high, intermediate, or low) will determine whether the probe identifies only naturally occurring sequences encoding the protein, alleles, or related sequences. Probes may also be used for the detection of related sequences, and should preferably contain at least 50% of the nucleotides from any of the protein-encoding sequences. The hybridisation probes of the subject invention may be DNA or RNA and are preferably derived from the nucleotide sequence of SEQ ID NO:4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, or 38, or from the genomic sequence including promoter, enhancer elements, and introns of the naturally occurring gene. Means for producing specific hybridisation probes for DNAs encoding the protein include the cloning of nucleic acid sequences encoding protein derivatives into vectors for the production of mRNA

- 40 -

probes. Such vectors are known in the art, commercially available, and may be used to synthesise RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labelled nucleotides. Hybridisation probes may be labelled by a variety of reporter groups, for example, radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding the protein may be used for the diagnosis of conditions or diseases, which are associated with expression of the protein. Examples of such conditions or diseases include, but are not limited to, pancreatic diseases and disorders, including diabetes. Polynucleotide sequences encoding the protein may also be used to monitor the progress of patients receiving treatment for pancreatic diseases and disorders, including diabetes. The polynucleotide sequences encoding the protein may be used in Southern or Northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; or in dip stick, pin, ELISA or chip assays utilising fluids or tissues from patient biopsies to detect altered gene expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding the protein may be useful in assays that detect activation or induction of various metabolic diseases and disorders, including obesity, diabetes, eating disorders, wasting syndromes (cachexia), pancreatic dysfunctions, arteriosclerosis, coronary artery disease (CAD), disorders related to ROS production, and neurodegenerative diseases. The nucleotide sequences encoding the protein may be labelled by standard methods, and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridisation complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. The presence of altered levels of target nucleotide sequences in the sample

- 41 -

indicates the presence of the associated disease. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or in monitoring the treatment of an individual patient.

5

In order to provide a basis for the diagnosis of disease associated with expression of the sequence of SEQ ID NO:4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, or 38, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, which encodes the protein, or a fragment thereof, under conditions suitable for hybridisation or amplification. Standard hybridisation may be quantified by comparing the values obtained from normal subjects with those from an experiment where a known amount of a substantially purified polynucleotide is used. Standard values obtained from normal samples may be compared with values obtained from samples from patients who are symptomatic for disease. Deviation between standard and subject values is used to establish the presence of disease. Once disease is established and a treatment protocol is initiated, hybridisation assays may be repeated on a regular basis to evaluate whether the level of expression in the patient begins to approximate that, which is observed in the normal patient. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

25

With respect to metabolic diseases and disorders, including obesity, diabetes, eating disorders, wasting syndromes (cachexia), pancreatic dysfunctions, arteriosclerosis, coronary artery disease (CAD), disorders related to ROS production, and neurodegenerative diseases presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual

30

- 42 -

clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the pancreatic diseases and disorders.

5

Additional diagnostic uses for oligonucleotides designed from the sequences of SEQ ID NO:4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, or 38 may involve the use of PCR. Such oligomers may be chemically synthesised, generated enzymatically, or produced from a recombinant source. Oligomers will preferably consist of two nucleotide sequences, one with sense orientation (5'.fwdarw.3') and another with antisense (3'.rarw.5'), employed under optimised conditions for identification of a specific gene or condition. The same two oligomers, nested sets of oligomers, or even a degenerate pool of oligomers may be employed under less stringent conditions for detection and/or quantification of closely related DNA or RNA sequences.

10  
15

Methods which may also be used to quantitate the gene expression include radiolabelling or biotinylating nucleotides, coamplification of a control nucleic acid, and standard curves onto which the experimental results are interpolated (Melby, P. C. et al. (1993) J. Immunol. Methods, 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantification of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantification.

20  
25

In another embodiment of the invention, the nucleic acid sequences, which encode the protein, may also be used to generate hybridisation probes, which are useful for mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome or to a specific region of the chromosome using well known techniques. Such techniques

30

- 43 -

include FISH, FACS, or artificial chromosome constructions, such as yeast artificial chromosomes, bacterial artificial chromosomes, bacterial P1 constructions or single chromosome cDNA libraries as reviewed in Price, C. M. (1993) *Blood Rev.* 7:127-134, and Trask, B. J. (1991) *Trends Genet.* 7:149-154. FISH (as described in Verma et al. (1988) *Human Chromosomes: A Manual of Basic Techniques*, Pergamon Press, New York, N.Y.) may be correlated with other physical chromosome mapping techniques and genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of *Science* (265:1981f). Correlation between the location of the gene encoding SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, or 39 on a physical chromosomal map and a specific disease, or predisposition to a specific disease, may help to delimit the region of DNA associated with that genetic disease.

The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier, or affected individuals. In situ hybridisation of chromosomal preparations and physical mapping techniques such as linkage analysis using established chromosomal markers may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms, or parts thereof, by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localised by genetic linkage to a particular genomic region, for example, AT to 11q22-23 (Gatti, R. A. et al. (1988) *Nature* 336:577-580), any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleotide sequences of the subject invention may also be used to detect

- 44 -

differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier, or affected individuals.

5 In another embodiment of the invention, the proteins, their catalytic or immunogenic fragments or oligopeptides thereof, an in vitro model, a genetically altered cell or animal, can be used for screening libraries of compounds in any of a variety of drug screening techniques. One can identify effectors, e.g. receptors, enzymes, proteins, ligands, or substrates that bind to, modulate or mimic the action of one or more of the proteins  
10 of the invention. The protein or fragment thereof employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between the protein and the agent tested, may be measured. Agents could also, either directly or indirectly, influence the activity of the proteins of  
15 the invention.

Candidate agents may also be found in kinase assays where a kinase substrate such as a protein or a peptide, which may or may not include modifications as further described below, or others are phosphorylated by  
20 the proteins or protein fragments of the invention. A therapeutic candidate agent may be identified by its ability to increase or decrease the enzymatic activity of the proteins of the invention. The kinase activity may be detected by change of the chemical, physical or immunological properties of the substrate due to phosphorylation.

25 One example could be the transfer of radioisotopically labelled phosphate groups from an appropriate donor molecule to the kinase substrate catalyzed by the polypeptides of the invention. The phosphorylation of the substrate may be followed by detection of the substrates autoradiography with techniques well known in the art.  
30

Yet in another example, the change of mass of the substrate due to its phosphorylation may be detected by mass spectrometry techniques.

One could also detect the phosphorylation status of a substrate with an analyte discriminating between the phosphorylated and unphosphorylated status of the substrate. Such an analyte may act by having different affinities for the phosphorylated and unphosphorylated forms of the substrate or by having specific affinity for phosphate groups. Such an analyte could be, but is not limited to an antibody or antibody derivative, a recombinant antibody-like structure, a protein, a nucleic acid, a molecule containing a complexed metal ion, an anion exchange chromatography matrix, an affinity chromatography matrix or any other molecule with phosphorylation dependent selectivity towards the substrate.

Such an analyte could be employed to detect the kinase substrate, which is immobilized on a solid support during or after an enzymatic reaction. If the analyte is an antibody, its binding to the substrate could be detected by a variety of techniques as they are described in Harlow and Lane, 1998, Antibodies, CSH Lab Press, NY. If the analyte molecule is not an antibody, it may be detected by virtue of its chemical, physical or immunological properties, being endogenously associated with it or engineered to it.

Yet in another example the kinase substrate may have features, designed or endogenous, to facilitate its binding or detection in order to generate a signal that is suitable for the analysis of the substrates phosphorylation status. These features may be, but are not limited to a biotin molecule or derivative thereof, a glutathione-S-transferase moiety, a moiety of six or more consecutive histidine residues, an amino acid sequence or hapten to function as an epitope tag, a fluorochrome, an enzyme or enzyme fragment. The kinase substrate may be linked to these or other features with a molecular spacer arm to avoid steric hindrance.

In one example the kinase substrate may be labelled with a fluorochrome. The binding of the analyte to the labelled substrate in solution may be followed by the technique of fluorescence polarization as it is described in the literature (see, for example, Deshpande, S. et al. (1999) Prog. Biomed.

- 46 -

Optics (SPIE) 3603:261; Parker, G. J. et al. (2000) J. Biomol. Screen. 5:77-88; Wu, P. et al. (1997) Anal. Biochem. 249:29-36). In a variation of this example, a fluorescent tracer molecule may compete with the substrate for the analyte to detect kinase activity by a technique which is known to those skilled in the art as indirect fluorescence polarization.

In vivo, the enzymatic kinase activity of the unmodified polypeptides of casein kinase delta and epsilon and dolichol kinase (CG8311 homologous protein) towards a substrate can be enhanced by appropriate stimuli, triggering the phosphorylation of casein kinase delta and epsilon and dolichol kinase. This may be induced in the natural context by extracellular or intracellular stimuli, such as signaling molecules or environmental influences. One may generate a system containing activated casein kinase delta and epsilon and dolichol kinase, may it be an organism, a tissue, a culture of cells or cell-free environment, by exogenously applying this stimulus or by mimicking this stimulus by a variety of the techniques, some of them described further below. A system containing activated casein kinase delta and epsilon and dolichol kinase may be produced (i) for the purpose of diagnosis, study, prevention, and treatment of diseases and disorders related to body-weight regulation and thermogenesis, for example, but not limited to, metabolic diseases such as obesity, as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, and gallstones.

In addition activity of Optic atrophy 1 protein (OPA1), cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta and epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, or formin-binding protein 21 against its physiological substrate(s) or derivatives thereof could be measured in cell-based assays. Agents may also interfere with

- 47 -

posttranslational modifications of the protein, such as phosphorylation and dephosphorylation, farnesylation, palmitoylation, acetylation, alkylation, ubiquitination, proteolytic processing, subcellular localization and degradation. Moreover, agents could influence the dimerization or  
5 oligomerization of the proteins of the invention or, in a heterologous manner, of the proteins of the invention with other proteins, for example, but not exclusively, docking proteins, enzymes, receptors, or translation factors. Agents could also act on the physical interaction of the proteins of this invention with other proteins, which are required for protein function,  
10 for example, but not exclusively, their downstream signaling.

Methods for determining protein-protein interaction are well known in the art. For example binding of a fluorescently labeled peptide derived from the interacting protein to the protein of the invention, or vice versa, could be  
15 detected by a change in polarisation. In case that both binding partners, which can be either the full length proteins as well as one binding partner as the full length protein and the other just represented as a peptide are fluorescently labeled, binding could be detected by fluorescence energy transfer (FRET) from one fluorophore to the other. In addition, a variety of  
20 commercially available assay principles suitable for detection of protein-protein interaction are well known in the art, for example but not exclusively AlphaScreen (PerkinElmer) or Scintillation Proximity Assays (SPA) by Amersham. Alternatively, the interaction of the proteins of the invention with cellular proteins could be the basis for a cell-based screening  
25 assay, in which both proteins are fluorescently labeled and interaction of both proteins is detected by analysing cotranslocation of both proteins with a cellular imaging reader, as has been developed for example, but not exclusively, by Cellomics or EvotecOAI. In all cases the two or more binding partners can be different proteins with one being the protein of the  
30 invention, or in case of dimerization and/or oligomerization the protein of the invention itself. Proteins of the invention, for which one target mechanism of interest, but not the only one, would be such protein/protein

- 48 -

interactions are Optic atrophy 1 protein (OPA1), cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta and epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, or formin-binding protein 21.

Of particular interest are screening assays for agents that have a low toxicity for mammalian cells. The term "agent" as used herein describes any molecule, e.g. protein or pharmaceutical, with the capability of altering or mimicking the physiological function of one or more of the proteins of the invention. Candidate agents encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 50 and less than about 2,500 Daltons. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise carbocyclic or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups.

Candidate agents are also found among biomolecules including peptides, saccharides, fatty acids, steroids, purines, pyrimidines, nucleic acids and derivatives, structural analogs or combinations thereof. Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides and oligopeptides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. Additionally, natural or synthetically produced libraries and

- 49 -

compounds are readily modified through conventional chemical, physical and biochemical means, and may be used to produce combinatorial libraries. Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification, etc. to produce structural analogs. Where the screening  
5 assay is a binding assay, one or more of the molecules may be joined to a label, where the label can directly or indirectly provide a detectable signal.

Another technique for drug screening, which may be used, provides for  
10 high throughput screening of compounds having suitable binding affinity to the protein of interest as described in published PCT application WO84/03564. In this method, as applied to the protein of the invention large numbers of different small test compounds are synthesised on a solid substrate, such as plastic pins or some other surface. The test compounds  
15 are reacted with the protein, or fragments thereof, and washed. Bound proteins are then detected by methods well known in the art. Purified proteins can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilise it on a solid  
20 support. In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding the protein specifically compete with a test compound for binding the protein. In this manner, the antibodies can be used to detect the presence of any peptide, which shares one or more antigenic determinants with the protein of the  
25 invention. In additional embodiments, the nucleotide sequences which encode the protein may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

30

The nucleic acids encoding the proteins of the invention can be used to generate transgenic cell lines and animals. These transgenic non-human

- 50 -

animals are useful in the study of the function and regulation of the proteins of the invention in vivo. Transgenic animals, particularly mammalian transgenic animals, can serve as a model system for the investigation of many developmental and cellular processes common to humans. A variety of non-human models of metabolic disorders can be used to test modulators of the protein of the invention. Misexpression (for example, overexpression or lack of expression) of the protein of the invention, particular feeding conditions, and/or administration of biologically active compounds can create models of metabolic disorders.

In one embodiment of the invention, such assays use mouse models of insulin resistance and/or diabetes, such as mice carrying gene knockouts in the leptin pathway (for example, ob (leptin) or db (leptin receptor) mice). Such mice develop typical symptoms of diabetes, show hepatic lipid accumulation and frequently have increased plasma lipid levels (see Bruning et al, 1998, Mol. Cell. 2:449-569). Susceptible wild type mice (for example C57Bl/6) show similar symptoms if fed a high fat diet. In addition to testing the expression of the proteins of the invention in such mouse strains (see EXAMPLES section), these mice could be used to test whether administration of a candidate modulator alters for example lipid accumulation in the liver, in plasma, or adipose tissues using standard assays well known in the art, such as FPLC, colorimetric assays, blood glucose level tests, insulin tolerance tests and others.

Transgenic animals may be made through homologous recombination in embryonic stem cells, where the normal locus of the gene encoding the protein of the invention is mutated. Alternatively, a nucleic acid construct encoding the protein is injected into oocytes and is randomly integrated into the genome. One may also express the genes of the invention or variants thereof in tissues where they are not normally expressed or at abnormal times of development. Furthermore, variants of the genes of the invention like specific constructs expressing anti-sense molecules or

- 51 -

expression of dominant negative mutations, which will block or alter the expression of the proteins of the invention may be randomly integrated into the genome. A detectable marker, such as lac Z or luciferase may be introduced into the locus of the genes of the invention, where upregulation  
5 of expression of the genes of the invention will result in an easily detectable change in phenotype. Vectors for stable integration include plasmids, retroviruses and other animal viruses, yeast artificial chromosomes (YACs), and the like. DNA constructs for homologous recombination will contain at least portions of the genes of the invention  
10 with the desired genetic modification, and will include regions of homology to the target locus. Conveniently, markers for positive and negative selection are included. DNA constructs for random integration do not need to contain regions of homology to mediate recombination. DNA constructs for random integration will consist of the nucleic acids encoding the  
15 proteins of the invention, a regulatory element (promoter), an intron and a poly-adenylation signal. Methods for generating cells having targeted gene modifications through homologous recombination are known in the field. For embryonic stem (ES) cells, an ES cell line may be employed, or embryonic cells may be obtained freshly from a host, e.g. mouse, rat,  
20 guinea pig, etc. Such cells are grown on an appropriate fibroblast-feeder layer and are grown in the presence of leukemia inhibiting factor (LIF). ES or embryonic cells may be transfected and can then be used to produce transgenic animals. After transfection, the ES cells are plated onto a feeder layer in an appropriate medium. Cells containing the construct may be  
25 selected by employing a selection medium. After sufficient time for colonies to grow, they are picked and analyzed for the occurrence of homologous recombination. Colonies that are positive may then be used for embryo manipulation and morula aggregation. Briefly, morulae are obtained from 4 to 6 week old superovulated females, the Zona Pellucida is removed  
30 and the morulae are put into small depressions of a tissue culture dish. The ES cells are trypsinized, and the modified cells are placed into the depression closely to the morulae. On the following day the aggregates are

- 52 -

transferred into the uterine horns of pseudopregnant females. Females are then allowed to go to term. Chimeric offsprings can be readily detected by a change in coat color and are subsequently screened for the transmission of the mutation into the next generation (F1-generation). Offspring of the F1-generation are screened for the presence of the modified gene and males and females having the modification are mated to produce homozygous progeny. If the gene alterations cause lethality at some point in development, tissues or organs can be maintained as allogenic or congenic grafts or transplants, or in vitro culture. The transgenic animals may be any non-human mammal, such as laboratory animal, domestic animals, etc., for example, mouse, rat, guinea pig, sheep, cow, pig, and others. The transgenic animals may be used in functional studies, drug screening, and other applications and are useful in the study of the function and regulation of the proteins of the invention in vivo.

Finally, the invention also relates to a kit comprising at least one of

- (a) an Optic atrophy 1 protein (OPA1), cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta and epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, or formin-binding protein 21 nucleic acid molecule or a fragment thereof;
- (b) an Optic atrophy 1 protein (OPA1), cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta and epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, or formin-binding protein 21 amino acid molecule or a fragment or an isoform thereof;
- (c) a vector comprising the nucleic acid of (a);
- (d) a host cell comprising the nucleic acid of (a) or the vector of (b);

- 53 -

- (e) a polypeptide encoded by the nucleic acid of (a);
- (f) a fusion polypeptide encoded by the nucleic acid of (a);
- (g) an antibody, an aptamer or another effector against the nucleic acid of (a) or the polypeptide of (b), (e) or (f) and
- 5 (h) an anti-sense oligonucleotide of the nucleic acid of (a).

The kit may be used for diagnostic or therapeutic purposes or for screening applications as described above. The kit may further contain user instructions.

10

## Figures

### Figure 1. Drosophila UCPy

- 15 Figure 1A. Full length cDNA sequence of Drosophila UCPy (SEQ ID NO:1)
- Figure 1B. Open reading frame of the deduced protein of Drosophila UCPy (SEQ ID NO:2).
- Figure 1C. Amino acid sequence of Drosophila UCPy (SEQ ID NO:3).

20

### Figure 2. The human homolog of CG8479

- Figure 2A. Blastn search result for CG8479
- Figure 2B. Predicted coding nucleotide sequence for the human homolog of CG8479 (SEQ ID NO:4)
- 25 Figure 2C. Predicted amino acid sequence for the human homolog of CG8479 (SEQ ID NO:5).

- Figure 3. Multiple Sequence alignment (ClustlIW 1.83) of Drosophila protein with Gadfly Accession Number CG8479 (referred to as CG8479 Dm), mouse (XP\_148016 Mm), and human (OPA1-5 Hs) homologs. The sequences are shown in the one letter code.
- 30

Figure 4. Expression of OPA1 in mammalian tissues

Figure 4A. Real time PCR analysis of OPA1 expression in wildtype mouse tissues.

5 Figure 5. The human homolog of CG5855 (cornichon)

Figure 5A. Blastn search result for CG5855

Figure 5B. Predicted coding nucleotide sequence for the human homolog (SEQ ID NO:6)

10 Figure 5C. Predicted amino acid sequence for the human homolog of CG5855 (SEQ ID NO:7).

Figure 6. Energy storage metabolites (ESM; triglyceride (TG) and glycogen) content of a cornichon (Gadfly Accession Number CG5855) mutant. Shown is the change of triglyceride content of HD-EP20292 flies caused by integration of the P-vector into the annotated transcription unit (column 3) in comparison to controls containing more than 2000 fly lines of the proprietary EP collection ('HD-control (TG)', column 1) and wildtype controls determined in more than 80 independent assays (referred to as 'WT-control (TG)' column 2). Also shown is the change of glycogen content of HD-EP20292 flies caused by integration of the P-vector into the annotated transcription unit (column 5) in comparison to controls (referred to as 'control (glycogen)' column 4).

Figure 7. The human homolog of CG1691 (Imp)

25 Figure 7A. Blastn search result for CG1691

Figure 7B. Predicted coding nucleotide sequence for the human homolog (SEQ ID NO:8)

Figure 7C. Predicted amino acid sequence for the human homolog of CG1691 (SEQ ID NO:9).

30

Figure 8. Human homolog of CG11988

Figure 8A. BlastP search result for CG11988 (neuralized)

- 55 -

Figure 8B. Predicted coding nucleotide sequence for the human homolog of CG11988; length of the sequence in base pairs (SEQ ID NO:10)

Figure 8C. Predicted amino acid sequence for the human homolog of CG11988; length of the sequence in amino acids (SEQ ID NO:11).

5

Figure 9. Expression of neuralized-like in mammalian tissues - Real time PCR analysis of neuralized-like expression in wildtype mouse tissues (DCT Pancreas = 23,34).

10 Figure 10. The human homolog of CG8311

Figure 10A. BlastP search result for CG8311

Figure 10B. Predicted coding sequence for the human homolog; length of the sequence in base pairs, referred to as bp. (SEQ ID NO:12)

15 Figure 10C. Predicted amino acid sequence for the human homolog of CG8311; length of the sequence in amino acids, referred to as aa. (SEQ ID NO:13)

Figure 10D. Transmembrane prediction for the human homolog protein.

20 Figure 11. Expression of CG8311 homolog in mammalian tissues - Real-time PCR analysis of the murine CG8311 homolog protein shows strongest expression in brown adipose tissue.

Figure 12. A human homolog of CG2048 (dco)

Figure 12A. BlastP search result for CG2048

25 Figure 12B. Predicted coding nucleotide sequence for the human homolog, Casein Kinase 1, delta; length of the sequence in base pairs, referred to as bp. (SEQ ID NO:14)

30 Figure 12C. Predicted amino acid sequence for the human homolog of CG2048; length of the sequence in amino acids, referred to as aa. (SEQ ID NO:15).

Figure 13. A human homolog of CG2048 (dco)

- 56 -

Figure 13A. BlastP search result for CG2048

Figure 13B. Predicted coding nucleotide sequence for the human homolog, Casein Kinase 1, epsilon; length of the sequence in base pairs, referred to as bp. (SEQ ID NO:16)

5 Figure 13C. Predicted amino acid sequence for the human homolog of CG2048; length of the sequence in amino acids, referred to as aa. (SEQ ID NO:17)

Figure 13D. ClustaW alignment of Drosophila GadFly Accession Number CG2048 (referred to as 'dCK I'), human casein kinase 1, delta (GenBank  
10 Accession Number NM\_001893.1; referred to as 'hCK I delta'), human casein kinase 1, epsilon (GenBank Accession Number XM\_009983.4; referred to as 'hCK I epsilon'), mouse casein kinase 1, delta (Accession Number AB028241.1; referred to as 'mCK I delta'), mouse casein kinase 1, epsilon (Accession Number NM\_013767.2; referred to as 'mCK I  
15 epsilon').

Figure 14. A human homolog of CG5320 (Gdh)

Figure 14A. BlastP search result for CG5320

Figure 14B. Predicted coding nucleotide sequence for the human homolog  
20 with Accession Number NM\_005271.1 (Glutamate dehydrogenase I); length of the sequence in base pairs, referred to as bp. (SEQ ID NO:18)

Figure 14C. Predicted amino acid sequence for the human homolog with Accession Number NM\_005271.1 (Glutamate dehydrogenase I); length of the sequence in amino acids, referred to as aa. (SEQ ID NO:19).

25

Figure 15. A human homolog of CG5320 (Gdh)

Figure 15A. BlastP search result for CG5320

Figure 15B. Predicted coding nucleotide sequence for the human homolog  
30 with Accession Number NT\_011746.5 (Glutamate dehydrogenase II); length of the sequence in base pairs, referred to as bp. (SEQ ID NO:20)

- 57 -

Figure 15C. Predicted amino acid sequence for the human homolog with Accession Number NT\_011746.5 (Glutamate dehydrogenase II); length of the sequence in amino acids, referred to as aa. (SEQ ID NO:21).

5 Figure 16. Energy storage metabolites (ESM; triglyceride (TG) and glycogen) content of a *Drosophila* Gdh (Gadfly Accession Number CG5320) mutant. Shown is the change of triglyceride content of HD-EP35207 flies caused by integration of the P-vector into the annotated transcription unit (column 3) in comparison to controls containing more  
10 than 2000 fly lines of the proprietary EP collection ('HD-control (TG)', column 1) and wildtype controls determined in more than 80 independent assays (referred to as 'WT-control (TG)' column 2). Also shown is the change of glycogen content of HD-EP35207 flies caused by integration of the P-vector into the annotated transcription unit (column 5) in comparison  
15 to controls (referred to as 'control (glycogen)' column 4).

Figure 17. Human homolog of CG3943 (kraken)

Figure 17A. tBlastN search result for CG3943

20 Figure 17B. Predicted coding nucleotide sequence for the human homolog of CG3943; length of the sequence in base pairs (SEQ ID NO:22)

Figure 17C. Predicted amino acid sequence for the human homolog of CG3943; length of the sequence in amino acids (SEQ ID NO:23)

Figure 17D. ClustalW alignment of *Drosophila* protein with GadFly Accession Number CG3943 (referred to as "drosophila") and the mouse  
25 (referred to as "mS0273353.1") and human (referred to as "HSC140179.1") homologs. The sequences are shown in the one-letter-code; shaded residues match exactly.

Figure 18. The human homolog of CG5216 (Sir2)

30 Figure 18A. Blastn search result for CG5216

Figure 18B. Predicted coding nucleotide sequence for the human homolog (SEQ ID NO:24)

- 58 -

Figure 18C. Predicted amino acid sequence for the human homolog of CG5216 (SEQ ID NO:25).

Figure 19. The human homolog of CG3758 (escargot)

5 Figure 19A. Blastn search result for CG3758

Figure 19B. Predicted coding nucleotide sequence for the human homolog (SEQ ID NO:26)

Figure 19C. Predicted amino acid sequence for the human homolog of CG3758 (SEQ ID NO:27).

10

Figure 20. Triglyceride content of *Drosophila* escargot (Gadfly Accession Number CG3758) mutants. Shown is the change of triglyceride content of HD-EP20506 (column 2), HD-EP20817 (column 3), and HD-EP26792 (column 4) flies caused by integration of the P-vector into the annotated  
15 transcription unit in comparison to controls containing all fly lines of the proprietary EP collection ('EP-control'), column 1).

Figure 21. The human homolog of CG3241 (msl-2)

Figure 21A. BlastP search result for CG3241

20 Figure 21B. Predicted coding nucleotide sequence for the human homolog with Accession Number AB046805.1 encoding hypothetical protein KIAA1585; length of the sequence in base pairs, referred to as bp. (SEQ ID NO:28)

25 Figure 21C. Predicted amino acid sequence for the human homolog of CG3241; length of the sequence in amino acids, referred to as aa. (SEQ ID NO:29)

30 Figure 21D. ClustaW alignment of *Drosophila* msl-2 (GadFly Accession Number CG3241; referred to as 'd Msl-2'), human msl-2 (GenBank Accession Number AB046805.1; referred to as 'hHIA1585'), and mouse msl-2 (GenBank Accession Number BF471233; referred to as 'mBF471233'). The sequences are shown in the one-letter-code; shaded residues match exactly.

- 59 -

Figure 22. Triglyceride content of a *Drosophila* msl-2 (Gadfly Accession Number CG3241) mutant. Shown is the change of triglyceride content of HD-EP25097 flies caused by integration of the P-vector into the annotated transcription unit (column 2) in comparison to controls containing all fly lines of the proprietary EP collection ('EP-control)', column 1).

Figure 23. The human homolog of CG11940 and triglyceride content of a *Drosophila* CG11940 mutant

Figure 23A. Blastn search result for CG11940

Figure 23B. Predicted coding nucleotide sequence for the human homolog (SEQ ID NO:30)

Figure 23C. Predicted amino acid sequence for the human homolog of CG11940 (SEQ ID NO:31)

Figure 23D. Triglyceride content of a *Drosophila* CG11940 (Gadfly Accession Number) mutant. Shown is the change of triglyceride content of HD-EP10934 flies caused by integration of the P-vector into the annotated transcription unit (column 2) in comparison to controls containing all fly lines of the proprietary EP collection ('EP-control)', column 1).

Figure 24. Human homolog of CG1624 (dappled)

Figure 24A. tBlastN search result for CG1624

Figure 24B. Predicted coding nucleotide sequence for the human homolog of CG1624; length of the sequence in base pairs (SEQ ID NO:32)

Figure 24C. Predicted amino acid sequence for the human homolog of CG1624; length of the sequence in amino acids (SEQ ID NO:33).

Figure 25. Human homolog of CG11753

Figure 25A. tBlastN search result for CG11753

Figure 25B. Predicted coding nucleotide sequence for the human homolog of CG11753; length of the sequence in base pairs (SEQ ID NO:34)

Figure 25C. Predicted amino acid sequence for the human homolog of CG11753; length of the sequence in amino acids (SEQ ID NO:35)

- 60 -

Figure 25D. ClustalW alignment of Drosophila protein with GadFly  
Accession Number CG11753 (referred to as "dCG11753") and the human  
(referred to as "hCG11753") and mouse (referred to as "mCG11753")  
homologs. The sequences are shown in the one-letter-code; shaded  
5 residues match exactly.

Figure 26. Human homolog of CG7262

Figure 26A. tBlastN search result for CG7262

Figure 26B. Predicted coding nucleotide sequence for the human homolog  
10 of CG7262; length of the sequence in base pairs (SEQ ID NO: 36)

Figure 26C. Predicted amino acid sequence for the human homolog of  
CG7262; length of the sequence in amino acids (SEQ ID NO: 37).

Figure 27. The human homolog of CG4291

15 Figure 27A. Blastn search result for CG4291

Figure 27B. Predicted coding nucleotide sequence for the human homolog  
(SEQ ID NO:38)

Figure 27C. Predicted amino acid sequence for the human homolog of  
CG4291 (SEQ ID NO:39).

20

The Examples illustrate the invention:

Example 1: Cloning of a Drosophila melanogaster gene with homology to  
25 human Uncoupling Proteins (UCPs)

A BLAST homology search was performed in a public database (NCBI/NIH)  
looking for Drosophila genes with sequence homology to the human UCP2  
and UCP3 genes. The search yielded sequence fragments of a family of  
30 Drosophila genes with UCP homology. They are clearly different to the  
next related mitochondrial proteins (oxoglutarate carrier).

- 61 -

Using the sequence fragment of one of this genes (here called dUCPy), a PCR primer pair was generated (Upper  
5' -CTAAACAAACAATTCCAAACATAG (SEQ ID NO: 40), Lower  
5' -AAAAGACATAGAAAATACGATAGT (SEQ ID NO: 41)) and a PCR  
5 reaction performed on Drosophila cDNA using standard PCR conditions.  
The amplification product was radioactively labeled and used to screen a  
cDNA library made from adult Drosophila flies (Stratagene). A full-length  
cDNA clone was isolated, sequenced and used for further experiments.  
The nucleotide sequence of UCPy is shown in SEQ ID NO:1 (see FIGURE  
10 1A), the coding sequence in SEQ ID NO:2 (see FIGURE 1B), and the  
deduced open reading frame is shown as SEQ ID NO:3 (see FIGURE 1C).

#### Example 2: Cloning of the dUCPy cDNA into an Drosophila expression vector

15 In order to test the effects of dUCPy expression in Drosophila cells the  
dUCPy cDNA was cloned into the expression vector pUAST (Ref.: Brand A  
& Perrimon N, Development 1993, 118:401-415) using the restriction sites  
NotI and KpnI. The resulting expression construct was injected into the  
20 germline of Drosophila embryos and Drosophila strains with a stable  
integration of the construct were generated. Since the expression vector  
pUAST is activated by the yeast transcription factor Gal4 which is normally  
absent from Drosophila cells dUCPy is not yet expressed in these  
transgenic animals. If pUAST-dUCPy flies are crossed with a second  
25 Drosophila strain that expresses Gal4 in a tissue specific manner the  
offspring flies of this mating will express dUCPy in the Gal4 expressing  
tissue.

30 The cross of pUAST-dUCPy flies with a strain that expresses Gal4 in all  
cells of the body (under control of the actin promoter) showed no viable  
offspring. This means that dUCPy overexpression in all body cells is lethal.

- 62 -

This finding is consistent with the assumption that dUCPy overexpression could lead to a collapse of the cellular energy production.

Expression of dUCPy in a non-vital organ like the eye (Gal4 under control of the eye-specific promoter of the "eyeless" gene) results in flies with visibly damaged eyes. This easily visible eye phenotype is the basis of a genetic screen for gene products that can modify UCP activity.

### Example 3: dUCPy modifier screen

Parts of the genomes of the strain with Gal4 expression in the eye and the strain carrying the pUAST-dUCPy construct were combined on one chromosome using genomic recombination. The resulting fly strain has eyes that are permanently damaged by dUCPy expression. Flies of this strain were crossed with flies of a large collection of mutagenized fly strains. In this mutant collection a special expression system (EP-element, Ref.: Rorth P, Proc Natl Acad Sci U S A 1996, 93(22):12418-12422) is integrated randomly in different genomic loci. The yeast transcription factor Gal4 can bind to the EP-element and activate the transcription of endogenous genes close the integration site of the EP-element. The activation of the genes therefore occurs in the same cells (eye) that overexpress dUCPy. Since the mutant collection contains several thousand strains with different integration sites of the EP-element it is possible to test a large number of genes whether their expression interacts with dUCPy activity. In case a gene acts as an enhancer of UCP activity the eye defect will be worsened; a suppressor will ameliorate the defect.

Using this screen genes with suppressing activity were discovered that were found to be the cornichon (GadFly Accession Number CG5855), neuralized (GadFly Accession Number CG11988), dco (GadFly Accession

- 63 -

Number CG2048), kraken (GadFly Accession Number CG3943), escargot (GadFly Accession Number CG3758), GadFly Accession Number CG11940, dappled (GadFly Accession Number CG1624), GadFly Accession Number CG11753, GadFly Accession Number CG7262, and GadFly Accession Number CG4291 genes in Drosophila. Using this screen genes with enhancing activity were discovered that was found to be the GadFly Accession Number CG8479, Imp (GadFly Accession Number CG1691), GadFly Accession Number CG8311, Gdh (GadFly Accession Number CG5320), Sir2 (GadFly Accession Number CG5216), and msl-2 (GadFly Accession Number CG3241) genes in Drosophila.

#### Example 4: Identification of human homologous genes and proteins

Genomic DNA neighbouring to the respective eye-defect rescuing EP-element was cloned by inverse PCR and sequenced. These sequences were used for BLAST searches in a public Drosophila gene database.

The database search indicated that the EP-element EP20761 which is enhancing the eye-phenotype is integrated in a predicted transcript annotated as CG8479 (Drosophila Genome Project), located on chromosome 2R, encoding for a protein with 65% homologies to human optic atrophy 1 protein (see FIGURE 2; SEQ ID NO: 4 and 5; GenBank Accession Number XP\_039926.2).

The database search indicated that the EP-element EP20292 which is suppressing the eye-phenotype is integrated in a predicted transcript annotated as FlyBase Symbol CG5855 (Drosophila Genome Project; gene *cni*), located on chromosome 2L, encoding for a protein with 76% homologies to human Cornichon-like protein (see FIGURE 5; SEQ ID NO:6 and 7; GenBank Accession Number NP\_005767.1).

- 64 -

The database search indicated that the EP-elements EP10858 and EP10570 which are enhancing the eye-phenotype are integrated in a predicted transcript annotated as CG1691 (Drosophila Genome Project), located on chromosome X, encoding for a protein with 63% homologies to human IGF-II mRNA binding protein 3 (see FIGURE 7; SEQ ID NO:8 and 9; 5 GenBank Accession Number XP\_004780.2).

The database search indicated that the EP-element EP31874 which is suppressing the eye-phenotype is integrated in a predicted transcript 10 annotated as CG11988 (Drosophila Genome Project; gene neur), located on chromosome 3R, encoding for a protein with 46% homology / 50% homology to human neuralized-like protein (see FIGURE 8; SEQ ID NO:10 and 11; GenBank Accession Number NM\_004210).

15 The database search indicated that the EP-element EP20700 which is enhancing the eye-phenotype is integrated in a predicted transcript annotated as CG8311 (Drosophila Genome Project), located on chromosome 2R, encoding for a protein with homologies to human KIAA1094 protein (GenBank Accession Number NM\_014908.1; SEQ ID 20 NO:12 and 13; see FIGURE 10); corresponding to patent WO0153486 (Sequence 69). Human KIAA1094 is 46% homologous and 29% identical to Drosophila CG8311 over 405 amino acids (see FIGURE 10A), and Human KIAA1094 is 50% homologous and 31% identical to Saccharomyces cerevisiae Sec59p (Accession Number NP\_013726.1) over 25 267 amino acids. The transmembrane prediction of the CG8311 homolog is shown in FIGURE 10D. The protein shows according to the THMM prediction program (Krogh et al., 2001, Journal of Molecular Biology 305(3):567-580; for example see <http://www.cbs.dtu.dk/services/TMHMM-2.0/>) 14 transmembrane domains, shown as black peaks in 30 FIGURE 10D. The human protein is most likely (74%) located in the plasma membrane, according to the publicly available prediction program PsortII (Horton and Nakai, 1996, Proc Int Conf Intell Syst Mol Biol. 4:109-15; for

- 65 -

example see <http://psort.nibb.ac.jp>). *Drosophila* CG8311 shows also homologies to mouse gene with the Accession Number AW553567.

The database search indicated that the EP-element EP31834 which is  
5 suppressing the eye-phenotype leads to the overexpression of a predicted transcript annotated as FlyBase Symbol CG2048 (*Drosophila* Genome Project), located on chromosome 3R, encoding for a protein with 93% homologies over 281 amino acids to human casein kinase delta (see  
FIGURE 12; SEQ ID NO:14 and 15; GenBank Accession Number  
10 NM\_001893.1; corresponding to patent US5846764 (Sequence 43), US5728806 (Sequence 43), and US5686412 (Sequence 34). CG2048 also shows high homologies to human casein kinase epsilon (see FIGURE 13; SEQ ID NO:16 and 17; GenBank Accession Number XM\_009983.4). *Drosophila* CG2048 shows also homologies to mouse genes with the  
15 Accession Numbers BAA88082 (murine casein kinase 1 delta), and NM\_013767 (murine casein kinase 1, epsilon). A Clusta-W alignment of *Drosophila* CG2048, both human homolog casein kinases, and the two homolog murine casein kinases was conducted and is shown in FIGURE 13D.

The database search indicated that the EP-element EP31710 which is  
20 enhancing the eye-phenotype is integrated in the promoter opposite to the driving direction of the predicted transcript annotated as CG5320 (*Drosophila* Genome Project), located on chromosome 3R, encoding for a  
25 protein with 78% homologies to 553 amino acids of human glutamate dehydrogenase GdH protein (GLUD1; see FIGURE 14; Seq ID NO:18 and 19; GenBank Accession Number NM\_005271.1 ); corresponding to patent WO0073801A2 (Sequence 453). CG5320 also shows high homologies (85% over 404 amino acids) to a second human glutamate dehydrogenase  
30 GdH protein (GLUD2; see FIGURE 15; Seq ID NO:20 and 21; GenBank Accession Number XP\_010438). *Drosophila* CG5320 shows also homologies to a mouse gene with the Accession Number NM\_008133.1.

- 66 -

The database search indicated that the EP-element EP20903 which is suppressing the eye-phenotype leads to the overexpression of a predicted transcript annotated as FlyBase Symbol CG3943 (Drosophila Genome Project; gene kraken), located on chromosome 2L, encoding for a protein with 54% homologies over 289 amino acids to a human hypothetical protein (see FIGURE 17; Seq ID NO:22 and 23; GenBank Accession Number CAC16804.1. A ClustalW alignment of Drosophila kraken and the mouse and human homologs was conducted and is shown in FIGURE 17D).

The database search indicated that the EP-element EP20105 which is enhancing the eye-phenotype is integrated in a predicted transcript annotated as CG5216 (Drosophila Genome Project; gene Sir2), located on chromosome 2L, encoding for a protein with 71% homologies to human Sirtuin protein (sirtuin 1; see FIGURE 18; SEQ ID NO:24 and 25; GenBank Accession Number XP\_008902.2).

The database search indicated that the EP-element EP20506 which is suppressing the eye-phenotype is integrated in a predicted transcript annotated as FlyBase Symbol CG3758 (Drosophila Genome Project; gene escargot), located on chromosome 2L, encoding for a protein with 85% homologies to human hypothetical protein, similar to Gonadotropin (see FIGURE 19; SEQ ID NO:26 and 27; GenBank Accession Number XP\_030528.1).

The database search indicated that the EP-element EP25097 which is enhancing the eye-phenotype is integrated in a predicted transcript annotated as FlyBase Symbol CG3241 (Drosophila Genome Project; gene msl-2), located on chromosome 2L, encoding for a protein with 58% homologies over 66 amino acids to human hypothetical protein KIAA1585 (see FIGURE 21; SEQ ID NO:28 and 29; GenBank Accession Number AB046805.1. A Clusta-W alignment of Drosophila msl-2 and the human

- 67 -

homolog was conducted and is shown in FIGURE 21D). Drosophila CG3241 shows also homologies to a mouse gene with the Accession Number BF471233.

5 The database search indicated that the EP-element EP11934 which is suppressing the eye-phenotype is integrated within the first (13kb) intron of a predicted transcript annotated as CG11940 (Drosophila Genome Project), located on chromosome X, encoding for a protein with 61% homologies to 226 amino acids of human alsin aslcr2 protein (see FIGURE  
10 23; Seq ID NO:30 and 31; GenBank Accession Number XP\_028059.1).

The database search indicated that the EP-element EP35393 which is suppressing the eye-phenotype is integrated in 3'-5' direction in a predicted transcript annotated as CG1624 (Drosophila Genome Project;  
15 gene dappled)), located on chromosome 3R, encoding for a protein with 68% homology to 171 amino acids, with 55% homology to 171 amino acids, and with 66% homology to 83 amino acids of a human protein (see FIGURE 24; Seq ID NO: 32 and 33; GenBank Accession Number XM\_067369).

20 The database search indicated that the EP-element EP32534 which is suppressing the eye-phenotype leads to the overexpression of a predicted transcript annotated as FlyBase Symbol CG11753 (Drosophila Genome Project), located on chromosome 3R, encoding for a protein with 61%  
25 homologies over 144 amino acids to a human protein (see FIGURE 25; SEQ ID NO:34 and 35; GenBank Accession Number XP\_029849.1). A ClustalW alignment of Drosophila CG11753 and the human and the mouse homolog was conducted and is shown in FIGURE 25D.

30 The database search indicated that the EP-element EP35056 which is suppressing the eye-phenotype is integrated in 3'-5' direction in the first intron of a predicted transcript annotated as CG7262 (Drosophila Genome

- 68 -

Project), located on chromosome 3R, encoding for a protein with homologies to human KIAA0095 protein (GenBank Accession Number NM\_014669; SEQ ID NO: 36 and 37; see FIGURE 26); corresponding to patent WO0018961 (Sequence 12). Human KIAA0095 is 55% homologous and 36% identical to Drosophila CG7262 over 823 amino acids (see FIGURE 26A). The protein shows according to the THMM prediction program (Krogh et al., 2001, Journal of Molecular Biology 305(3):567-580; for example see <http://www.cbs.dtu.dk/services/TMHMM-2.0/>) no transmembrane domains. The human protein is most likely (52%) located in the plasma membrane, according to the publicly available prediction program PsortII (Horton and Nakai, 1996, Proc Int Conf Intell Syst Mol Biol. 4:109-15; for example see <http://psort.nibb.ac.jp>).

The database search indicated that the EP-element EP20903 which is suppressing the eye-phenotype is integrated in a predicted transcript annotated as FlyBase Symbol CG4291 (Drosophila Genome Project), located on chromosome 2L, encoding for a protein with 45% homologies to human formin binding protein 21 (FBP21; see FIGURE 27; SEQ ID NO:38 and 39; GenBank Accession Number XP\_049375.1).

#### Example 5: Measurement of energy storage metabolites (ESM) content

Mutant flies are obtained from a fly mutation stock collection. The flies are grown under standard conditions known to those skilled in the art. In the course of the experiment, additional feedings with bakers yeast (*Saccharomyces cerevisiae*) are provided for the EP-lines HD-EP20292, HD-35207, HD-EP20506, HD-EP20817, HD-EP26792, HD-EP25097, and HD-EP10934. The average change of triglyceride and glycogen (herein referred to as energy storage metabolites, ESM) content of Drosophila containing the EP-vector as homozygous or hemizygous viable integration was investigated in comparison to control flies, respectively (see FIGURES 6, 16, 20, 22, and 23D). For determination of ESM content, flies were

- 69 -

incubated for 5 min at 90°C in an aqueous buffer using a waterbath, followed by hot extraction. After another 5 min incubation at 90°C and mild centrifugation, the triglyceride content of the flies extract was determined using Sigma Triglyceride (INT 336-10 or -20) assay by measuring changes in the optical density according to the manufacturer's protocol, and the glycogen content of the flies extract was determined using Roche (Starch UV-method Cat. No. 0207748) assay by measuring changes in the optical density according to the manufacturer's protocol. As a reference the protein content of the same extract was measured using BIO-RAD DC Protein Assay according to the manufacturer's protocol. These experiments and assays were repeated several times.

The average triglyceride level ( $\mu\text{g}$  triglyceride/ $\mu\text{g}$  protein) of all flies of the EP collection (referred to as 'EP-control') is shown as 100% in the first column in FIGURES 20, 22, and 23D. The average triglyceride level ( $\mu\text{g}$  triglyceride/ $\mu\text{g}$  protein) of 2108 fly lines of the proprietary EP-collection (referred to as 'HD-control (TG)') is shown as 100% in the first column in FIGURES 6 and 16. The average triglyceride level ( $\mu\text{g}$  triglyceride/ $\mu\text{g}$  protein) of *Drosophila* wildtype strain Oregon R flies determined in 84 independent assays (referred to as 'WT-control (TG)') is shown as 102% in the second column in FIGURES 6 and 16. The average glycogen level ( $\mu\text{g}$  glycogen/ $\mu\text{g}$  protein) of an internal assay control consisting of two different wildtype strains and an inconspicuous EP-line of the HD stock collection (referred to as 'control (glycogen)') is shown as 100% in the fourth column in FIGURES 6 and 16. Standard deviations of the measurements are shown as thin bars.

HD-EP20292 homozygous flies show constantly a lower triglyceride content ( $\mu\text{g}$  triglyceride/ $\mu\text{g}$  protein) than the controls (column 3 in FIGURE 6, 'HD-EP20292 (TG)'). HD-EP20292 homozygous flies also show a lower glycogen content ( $\mu\text{g}$  glycogen/ $\mu\text{g}$  protein) than the controls (column 5 in FIGURE 6, 'HD-EP20292 (glycogen)'). Therefore, the loss of gene activity

- 70 -

is responsible for changes in the metabolism of the energy storage metabolites.

HD-35207 homozygous flies show constantly a lower triglyceride content ( $\mu\text{g}$  triglyceride/ $\mu\text{g}$  protein) than the controls (column 3 in FIGURE 16, 'HD-35207 (TG)'). HD-35207 homozygous flies also show a lower glycogen content ( $\mu\text{g}$  glycogen/ $\mu\text{g}$  protein) than the controls (column 5 in FIGURE 16, 'HD-35207 (glycogen)'). Therefore, the loss of gene activity is responsible for changes in the metabolism of the energy storage metabolites.

HD-EP20506, HD-EP20817, and HD-EP26792 homozygous flies show constantly a higher triglyceride content ( $\mu\text{g}$  triglyceride/ $\mu\text{g}$  protein) than the controls (column 2 in FIGURE 20, 'HD-EP20506'; column 3 in FIGURE 20 'HD-EP20817', and column 4 in FIGURE 20, 'HD-EP26792'). Therefore, the loss of gene activity is responsible for changes in the metabolism of the energy storage triglycerides.

HD-EP25097 homozygous flies show constantly a higher triglyceride content ( $\mu\text{g}$  triglyceride/ $\mu\text{g}$  protein) than the controls (column 2 in FIGURE 22, 'HD-EP25097'). Therefore, the loss of gene activity is responsible for changes in the metabolism of the energy storage triglycerides.

HD-EP10934 hemizygous flies show constantly a higher triglyceride content ( $\mu\text{g}$  triglyceride/ $\mu\text{g}$  protein) than the controls (column 3 in FIGURE 23D, 'HD-EP10934'). Therefore, the loss of gene activity is responsible for changes in the metabolism of the energy storage triglycerides.

#### Example 6: Expression profiling experiments

To analyze the expression of the polypeptides disclosed in this invention in mammalian tissues, several mouse strains (preferably mice strains

- 71 -

C57Bl/6J, C57Bl/6 ob/ob, and C57Bl/KS db/db which are standard model systems in obesity and diabetes research) were purchased from Harlan Winkelmann (33178 Borcheln, Germany) and maintained under constant temperature (preferably 22°C), 40 percent humidity and a light / dark cycle of preferably 14 / 10 hours. The mice were fed a standard diet (for example, from ssniff Spezialitäten GmbH, order number sniff M-Z V1126-000). For the fasting experiment ("fasted wild type mice"), wild type mice were starved for 48 h without food, but only water supplied ad libitum (see, for example, Schnetzler et al. J Clin Invest 1993 Jul;92(1):272-80, Mizuno et al. Proc Natl Acad Sci U S A 1996 Apr 16;93(8):3434-8). Animals were sacrificed at an age of 6 to 8 weeks. The animal tissues were isolated according to standard procedures known to those skilled in the art, snap frozen in liquid nitrogen and stored at -80°C until needed.

For analyzing the role of the proteins disclosed in this invention in the in vitro differentiation of different mammalian cell culture cells for the conversion of pre-adipocytes to adipocytes, mammalian fibroblast (3T3-L1) cells (e.g., Green & Kehinde, Cell 1: 113-116, 1974) were obtained from the American Tissue Culture Collection (ATCC, Hanassas, VA, USA; ATCC- CL 173). 3T3-L1 cells were maintained as fibroblasts and differentiated into adipocytes as described in the prior art (e.g., Qiu. et al., J. Biol. Chem. 276:11988-95, 2001; Slieker et al., BBRC 251: 225-9, 1998). At various time points of the differentiation procedure, beginning with day 0 (day of confluence) and day 2 (hormone addition; for example, dexamethasone and 3-isobutyl-1-methylxanthine), up to 10 days of differentiation, suitable aliquots of cells were taken every two days. Alternatively, mammalian fibroblast 3T3-F442A cells (e.g., Green & Kehinde, Cell 7: 105-113, 1976) were obtained from the Harvard Medical School, Department of Cell Biology (Boston, MA, USA). 3T3-F442A cells were maintained as fibroblasts and differentiated into adipocytes as described previously (Djian, P. et al., J. Cell. Physiol., 124:554-556,

- 72 -

1985). At various time points of the differentiation procedure, beginning with day 0 (day of confluence and hormone addition, for example, Insulin), up to 10 days of differentiation, suitable aliquots of cells were taken every two days. 3T3-F442A cells are differentiating in vitro already in the  
5 confluent stage after hormone (insulin) addition.

RNA was isolated from mouse tissues or cell culture cells using Trizol Reagent (e.g. from Invitrogen, Karlsruhe, Germany) and further purified with the RNeasy Kit (for example, from Qiagen, Germany) in combination  
10 with a DNase-treatment according to the instructions of the manufacturers and as known to those skilled in the art. Total RNA was reverse transcribed (Superscript II RNaseH- Reverse Transcriptase, e.g. from Invitrogen, Germany) and subjected to Taqman analysis using the Taqman 2xPCR Master Mix (e.g. from Applied Biosystems, Weiterstadt, Germany;  
15 the Mix contains according to the Manufacturer for example AmpliTaq Gold DNA Polymerase, AmpErase UNG, dNTPs with dUTP, passive reference Rox and optimized buffer components) on a GeneAmp 5700 Sequence Detection System (e.g. from Applied Biosystems, Weiterstadt, Germany).

20 The Taqman analysis of the CG8479 homologous protein (OPA1) was performed using the following primer/probe pair: mouse OPA1 forward primer (SEQ ID NO: 42): 5'- GCC TGG GAG ACT CTA CAA GAG G -3'; mouse OPA1 reverse primer (SEQ ID NO: 43): 5'- AAT ATG TCG TCG TGT TCC TTT CC -3'; Taqman probe (SEQ ID NO: 44): (5/6-FAM) (5/6-FAM)  
25 TTT CCC GCT TCA TGA CAG AAC CCA A (5/6-TAMRA).

The Taqman analysis of the neuralized homologous protein was performed using the following primer/probe pair: mouse neuralized forward primer (SEQ ID NO: 45): 5'- TCA AGG ACA TCA TCA AGA CCT ACC-3'; mouse  
30 neuralized reverse primer (SEQ ID NO: 46): 5prime- GGG AGA CGT TGT GCA GGT G -3'; Taqman probe (FAM/TAMRA) (SEQ ID NO: 47): 5'- CAG CTC CTA GCC CAC TGC AGA GCC -3'.

- 73 -

The Taqman analysis of the CG8311 homologous protein was performed using the following primer/probe pair: mouse forward primer (SEQ ID NO: 48): 5'-GGAGGCCACAGTATCACCCA-3'; mouse reverse primer (SEQ ID NO: 49): 5'-AAGGAGCAAGAGCCCTGGTC-3'; Taqman probe (FAM/TAMRA) (SEQ ID NO: 50): 5'-ACCCACAGCCAAGACCCCAGCA-3'.

As shown in Figure 4, real time PCR (Taqman) analysis of the expression of the OPA-1 RNA in mammalian (mouse) tissues revealed that OPA-1 is expressed in different mammalian tissues, showing 2 to 3 fold higher levels of expression in BAT, hypothalamus, brain, muscle and heart when compared to other tissues. BAT, brain, muscle and heart represent tissues with the major catabolic activity in the body. The high expression levels of OPA-1 in these tissues indicate, that OPA-1 is involved in the metabolism of tissues relevant for the metabolic syndrome.

As shown in Figure 9, real time PCR (Taqman) analysis of the expression of the neuralized RNA in mammalian (mouse) tissues revealed that neuralized is highly expressed in muscle, hypothalamus, brain and testis. The high expression levels in muscle when compared to other tissues is indicative for a role in the metabolism in one of the major catabolic tissues of the body.

The Taqman analysis revealed that transcript levels of the CG8311 homologous protein show a prominent peak in brown adipose tissue compared to several other mouse tissues and organs. In Figure 11, Rel. RNA refers to relative RNA expression in the corresponding tissue, expressed as levels in percent [%]. The pancreas tissue was set as reference level to zero. The mouse tissue tested are shown on the vertical line; BAT refers to brown adipose tissue; WAT refers to white adipose tissue.

- 74 -

Example 7: In vitro assays for the determination of triglyceride and glycogen storage

Obesity is known to be caused by different reasons such as non-insulin  
5 dependent diabetes, increase in triglycerides, increase in carbohydrate  
bound energy and low energy expenditure. For example, an increase in  
energy expenditure (and thus, lowering the body weight) would include the  
elevated utilization of both circulating and intracellular glucose and  
triglycerides, free or stored as glycogen or lipids as fuel for energy and/or  
10 heat production. In this invention, we therefore show the cellular level of  
triglycerides and glycogen in cells overexpressing the protein of the  
invention.

Retroviral infection of preadipocytes

15 Packaging cells were transfected with retroviral plasmids pLPCX carrying  
the mouse transgene encoding a protein of the invention and a selection  
marker using calcium phosphate procedure. Control cells were infected  
with pLPCX carrying no transgene. Briefly, exponentially growing  
packaging cells were seeded at a density of 350,000 cells per 6-well in 2  
20 ml DMEM + 10 % FCS one day before transfection. 10 min before  
transfection chloroquine was added directly to the overlying medium (25  
 $\mu$ M end concentration). A 250  $\mu$ l transfection mix consisting of 5  $\mu$ g  
plasmid-DNA (candidate:helper-virus in a 1:1 ratio) and 250 mM  $\text{CaCl}_2$  was  
prepared in a 15 ml plastic tube. The same volume of 2 x HBS (280  $\mu$ M  
25 NaCl, 50  $\mu$ M HEPES, 1.5 mM  $\text{Na}_2\text{HPO}_4$ , pH 7.06) was added and air  
bubbles were injected into the mixture for 15 sec. The transfection mix  
was added drop wise to the packaging cells, distributed and the cells were  
incubated at 37°C, 5 %  $\text{CO}_2$  for 6 hours. The cells were washed with PBS  
and the medium was exchanged with 2 ml DMEM + 10 % CS per 6-well.  
30 One day after transfection the cells were washed again and incubated for  
2 days of virus collection in 1 ml DMEM + 10 % CS per 6-well at 32°C,  
5 %  $\text{CO}_2$ . The supernatant was then filtered through a 0.45  $\mu$ m cellulose

- 75 -

acetate filter and polybrene (end concentration 8  $\mu\text{g/ml}$ ) was added. Mammalian fibroblast (3T3-L1) cells in a sub-confluent state were overlaid with the prepared virus containing medium. The infected cells were selected for 1 week with 2  $\mu\text{g/ml}$  puromycin. Following selection the cells  
5 were checked for transgene expression by western blot and immunofluorescence. Over expressing cells were seeded for differentiation.

3T3-L1 cells were maintained as fibroblasts and differentiated into adipocytes as described in the prior art and supra. For analysing the role of  
10 the proteins disclosed in this invention in the in vitro assays for the determination of triglyceride storage, synthesis and transport were performed.

#### Preparation of cell lysates for analysis of metabolites

15 Starting at confluence (d0), cell media was changed every 48 hours. Cells and media were harvested 8 hours prior to media change as follows. Media was collected, and cells were washed twice in PBS prior to lyses in 600  $\mu\text{l}$  HB-buffer (0.5% polyoxyethylene 10 tridecylethane, 1 mM EDTA, 0.01M  $\text{NaH}_2\text{PO}_4$ , pH 7.4). After inactivation at 70°C for 5 minutes, cell lysates  
20 were prepared on Bio 101 systems lysing matrix B (0.1 mm silica beads; Q-Biogene, Carlsbad, USA) by agitation for 2 x 45 seconds at a speed of 4.5 (Fastprep FP120, Bio 101 Thermosavant, Holbrook, USA). Supernatants of lysed cells were collected after centrifugation at 3000 rpm for 2 minutes, and stored in aliquots for later analysis at -80°C.

#### Changes in cellular triglyceride levels during adipogenesis

Cell lysates and media were simultaneously analysed in 96-well plates for total protein and triglyceride content using the Bio-Rad DC Protein assay reagent (Bio-Rad, Munich, Germany) according to the manufacturer's  
30 instructions and a modified enzymatic triglyceride kit (GPO-Trinder; Sigma) briefly final volumes of reagents were adjusted to the 96-well format as follows: 10  $\mu\text{l}$  samples were incubated with 200  $\mu\text{l}$  reagent A for 5 minutes

- 76 -

at 37°C. After determination of glycerol (initial absorbance at 540 nm), 50  $\mu$ l reagent B was added followed by another incubation for 5 minutes at 37°C (final absorbance at 540 nm). Glycerol and triglyceride concentrations were calculated using a glycerol standard set (Sigma) for the standard curve included in each assay.

#### Changes in cellular glycogen levels during adipogenesis

Cell lysates and media were simultaneously analysed in triplicates in 96-well plates for total protein and glycogen content using the Bio-Rad DC Protein assay reagent (Bio-Rad, Munich, Germany) according to the manufacturer's instructions and an enzymatic starch kit from Hoffmann-La Roche (Basel, Switzerland). 10- $\mu$ l samples were incubated with 20- $\mu$ l amyloglucosidase solution for 15 minutes at 60°C to digest glycogen to glucose. The glucose is further metabolised with 100  $\mu$ l distilled water and 100  $\mu$ l of enzyme cofactor buffer and 12  $\mu$ l of enzyme buffer (hexokinase and glucose phosphate dehydrogenase). Background glucose levels are determined by subtracting values from a duplicate plate without the amyloglucosidase. Final absorbance is determined at 340 nm. HB-buffer as blank, and a standard curve of glycogen (Hoffmann-La Roche) were included in each assay. Glycogen content in samples were calculated using a standard curve.

#### Synthesis of lipids during adipogenesis

During the terminal stage of adipogenesis (day 12) cells were analysed for their ability to metabolise lipids. A modified protocol to the method of Jensen et al (2000) for lipid synthesis was established. Cells were washed 3 times with PBS prior to serum starvation in Krebs-Ringer-Bicarbonate-Hepes buffer (KRBH; 134 mM NaCl, 3.5 mM KCl, 1.2 mM  $\text{KH}_2\text{PO}_4$ , 0.5 mM  $\text{MgSO}_4$ , 1.5 mM  $\text{CaCl}_2$ , 5 mM  $\text{NaHCO}_3$ , 10 mM Hepes, pH 7.4), supplemented with 0.1% FCS for 2.5h at 37°C. For insulin-stimulated lipid synthesis, cells were incubated with 1  $\mu$ M bovine

- 77 -

insulin (Sigma; carrier: 0.005N HCl) for 45 min at 37°C. Basal lipid synthesis was determined with carrier only.  $^{14}\text{C}(\text{U})$ -D-Glucose (NEN Life Sciences) in a final activity of  $1\mu\text{Ci/Well/ml}$  in the presence of 5 mM glucose was added for 30 min at 37°C. For the calculation of background radioactivity, 25  $\mu\text{M}$  Cytochalasin B (Sigma) was used. All assays were performed in duplicate wells. To terminate the reaction, cells were washed 3 times with ice cold PBS, and lysed in 1 ml 0.1N NaOH. Protein concentration of each well was assessed using the standard Biuret method (Protein assay reagent; Bio-Rad). Total lipids were separated from aqueous phase after overnight extraction in Insta-Fluor scintillation cocktail (Packard Bioscience) followed by scintillation counting.

#### Transport and metabolism of free fatty acids during adipogenesis

During the terminal stage of adipogenesis (d12) cells were analysed for their ability to transport long chain fatty acid across the plasma membrane. A modified protocol to the method of Abumrad et al (1991) (Proc. Natl. Acad. Sci. USA, 1991: 88; 6008-12) for cellular transportation of fatty acid was established. In summary, cells were washed 3 times with PBS prior to serum starvation. This was followed by incubation in KRBH buffer supplemented with 0.1 % FCS for 2.5h at 37°C. Uptake of exogenous free fatty acids was initiated by the addition of isotopic media containing non radioactive oleate and  $(^3\text{H})$ oleate (NEN Life Sciences) complexed to serum albumin in a final activity of  $1\mu\text{Ci/Well/ml}$  in the presence of 5 mM glucose for 30min at room temperature (RT). For the calculation of passive diffusion (PD) in the absence of active transport (AT) across the plasma membrane 20mM of phloretin in glucose free media (Sigma) was added for 30 min at RT. All assays were performed in duplicate wells. To terminate the active transport 20mM of phloretin in glucose free media was added to the cells. Cells were lysed in 1 ml 0.1N NaOH and the protein concentration of each well were assessed using the standard Biuret method (Protein assay reagent; Bio-Rad). Esterified fatty acids were

- 78 -

separated from free fatty acids using overnight extraction in Insta-Fluor scintillation cocktail (Packard Bioscience) followed by scintillation counting.

#### Example 8: Glucose uptake assay

5

For the determination of glucose uptake, cells were washed 3 times with PBS prior to serum starvation in KRBH buffer supplemented with 0.1 % FCS and 0.5mM glucose for 2.5h at 37°C. For insulin-stimulated glucose uptake, cells were incubated with 1  $\mu$ M bovine insulin (Sigma; carrier: 10 0.005N HCl) for 45 min at 37°C. Basal glucose uptake was determined with carrier only. Non-metabolizable 2-deoxy-<sup>3</sup>H-D-glucose (NEN Life Science, Boston, USA) in a final activity of 0,4  $\mu$ Ci/Well/ml was added for 30 min at 37°C. For the calculation of background radioactivity, 25  $\mu$ M cytochalasin B (Sigma) was used. All assays were performed in duplicate 15 wells. To terminate the reaction, cells were washed 3 times with ice cold PBS, and lysed in 1 ml 0.1N NaOH. Protein concentration of each well was assessed using the standard Biuret method (Protein assay reagent; Bio-Rad), and scintillation counting of cell lysates in 10 volumes Ultima-gold cocktail (Packard Bioscience, Groningen, Netherlands) was 20 performed.

#### Example 9: Generation and analysis of transgenic mice

##### Generation of the transgenic animals

25

Mouse cDNA encoding OPA1, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta and epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, or formin-binding protein 21, was isolated 30 from mouse brown adipose tissue (BAT) using standard protocols as known to those skilled in the art. The cDNA was amplified by RT-PCR and point mutations were introduced into the cDNA.

- 79 -

The resulting mutated cDNA was cloned into a suitable transgenic expression vector. The transgene was microinjected into the male pronucleus of fertilized mouse embryos (preferably strain C57/BL6/CBA F1 (Harlan Winkelmann)). Injected embryos were transferred into  
5 pseudo-pregnant foster mice. Transgenic founders were detected by PCR analysis. Two independent transgenic mouse lines containing the construct were established and kept on a C57/BL6 background. Briefly, founder animals were backcrossed with C57/BL6 mice to generate F1 mice for analysis. Transgenic mice were continuously bred onto the C57/BL6  
10 background. The expression of the proteins of the invention can be analyzed by taqman analysis as described above, and further analysis of the mice can be done as known to those skilled in the art.

All publications and patents mentioned in the above specification are herein  
15 incorporated by reference.

Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention  
20 has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the  
25 scope of the following claims.

- 80 -

### Claims

1. A pharmaceutical composition comprising a nucleic acid molecule of  
5 the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding  
protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or  
epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1,  
escargot homolog, KIAA1585 protein, CG11940 homolog, dappled  
homolog, CG11753 homolog, KIAA0095 protein, and/or  
10 formin-binding protein 21 gene family or a polypeptide encoded  
thereby or a fragment or a variant of said nucleic acid molecule or  
said polypeptide or an antibody, an aptamer or another receptor  
recognizing said nucleic acid molecule or a said polypeptide encoded  
thereby together with pharmaceutically acceptable carriers, and/or  
15 diluents and/or adjuvants.
2. The composition of claim 1, wherein the nucleic acid molecule is a  
vertebrate or insect nucleic acid, particularly a human nucleic acid as  
shown in SEQ ID NO:4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26,  
20 28, 30, 32, 34, 36, and/or 38 or a nucleic acid having a nucleotide  
sequence complementary thereto or a fragment or a variant thereof.
3. The composition of claim 1 or 2, wherein said nucleic acid molecule  
(a) hybridizes at 50°C in a solution containing 1 x SSC and 0.1%  
25 SDS to a nucleic acid molecule encoding an amino acid  
sequence of SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23,  
25, 27, 29, 31, 33, 35, 37, or 39; and/or a nucleic acid  
molecule complementary thereto;  
(b) is degenerate with respect to the nucleic acid molecule of (a);  
30 (c) encodes a polypeptide which is at least 85%, preferably at  
least 90%, more preferably at least 95%, more preferably at  
least 98% and up to 99,6% identical to a protein as shown in

- 81 -

SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29,  
31, 33, 35, 37, or 39;

- (d) differs from the nucleic acid molecule of (a) to (g) by mutation  
and wherein said mutation causes an alteration, deletion,  
duplication or premature stop in the encoded polypeptide; or  
(e) comprises a partial sequence of any of the nucleotide  
sequences of (a) to (d) having a length of at least 15 bases.

4. The composition of any one of claims 1-3, wherein the nucleic acid  
molecule is a DNA molecule, particularly a cDNA or a genomic DNA.

5. The composition of any one of claims 1-4, wherein said nucleic acid  
encodes a polypeptide contributing to regulating the energy  
homeostasis and/or to membrane stability and/ or function in  
organelles such as mitochondria and/or peroxisomes.

6. The composition of claim 5, wherein said polypeptide participates in  
the maintenance of said membrane.

7. The composition of any one of claims 1 to 6, wherein said  
polypeptide is a transporter molecule and/or a regulator of a  
transporter molecule.

8. The composition of any one of claims 1 to 7, wherein said  
polypeptide is a modifying polypeptide.

9. The composition of claim 8, wherein said modifying polypeptide is a  
modifier of mitochondrial proteins.

10. The composition of claim 9, wherein said mitochondrial protein is a  
member of the UCP family.

- 82 -

11. The composition of claim 10, wherein said member of the UCP family is UCP1, UCP2, UCP3, UCP4, UCP5, StUCP or AtUCP.

5 12. The composition of any one of claims 1-11, wherein said nucleic acid molecule is a recombinant nucleic acid molecule.

13. The composition of any one of claims 1-12, wherein the nucleic acid molecule is a vector, particularly an expression vector.

10 14. The composition of any one of claims 1-11, wherein the polypeptide is a recombinant polypeptide.

15 15. The composition of claim 14, wherein said recombinant polypeptide is a fusion polypeptide.

16. The composition of any one of claims 1-11, wherein said nucleic acid molecule is selected from hybridization probes, primers and anti-sense oligonucleotides.

20 17. The composition of any one of claims 1-16 which is a diagnostic composition.

18. The composition of any one of claims 1-17 which is a pharmaceutical composition.

25 19. The composition of any one of claims 1-18 for the manufacture of an agent for detecting and/or verifying, for the treatment, alleviation and/or prevention of diseases and disorders related to body-weight regulation and thermogenesis, for example, but not limited to, metabolic diseases such as obesity, as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia,

30

- 83 -

osteoarthritis and gallstones and disorders related to ROS defence, such as diabetes mellitus, neurodegenerative disorders, mitochondrial disorders and others, in cells, cell masses, organs and/or subjects.

5

20. A vector comprising a nucleic acid molecule of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or formin-binding protein 21 gene family operatively linked to an expression control sequence.

10

15

21. A host transformed with the vector of claim 20.

22. A method for producing a polypeptide comprising culturing the host of claim 21 under suitable conditions and isolating the polypeptide produced.

20

23. An antibody, fragment or derivative thereof or an aptamer or another receptor specifically recognizing a nucleic acid molecule of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or formin-binding protein 21 gene family or a polypeptide encoded thereby.

25

30

24. An anti-sense oligonucleotide, primer or hybridization probe for a nucleic acid molecule of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein,

- 84 -

casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or formin-binding protein 21 gene family.

5

25. A non-human transgenic animal expressing a polypeptide of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or formin-binding protein 21 family, which is transfected with the vector of claim 20.

10

26. A non-human transgenic animal, wherein expression of a nucleic acid molecule of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or formin-binding protein 21 gene family or a homolog, paralog or ortholog thereof is silenced and/or mutated.

20

27. The non-human animal of claim 25 or 26 which is selected from the group consisting of mouse, rat, sheep, hamster, pig, dog, monkey, rabbit, calf, horse, nematodes, fly and fish.

25

28. Use of a nucleic acid molecule of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753

30

- 85 -

homolog, KIAA0095 protein, and/or formin-binding protein 21 gene family or a polypeptide encoded thereby or a fragment or a variant of said nucleic acid molecule or said polypeptide or an antibody, an aptamer or another receptor recognizing said nucleic acid molecule or a polypeptide encoded thereby for controlling the function of a gene and/or a gene product which is influenced and/or modified by said polypeptide.

29. The use of claim 28, wherein said gene and/or gene product is a gene and/or gene product expressed in organelles.

30. The use of claim 29, wherein said organelle is a mitochondrion or a peroxisome.

31. The use of any one of claims 28 to 30, wherein said gene and/or gene product is a member of the UCP family.

32. Use of a nucleic acid molecule of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or formin-binding protein 21 gene family or a polypeptide encoded thereby or a fragment or a variant of said nucleic acid molecule or said polypeptide or an antibody, an aptamer or another receptor recognizing said nucleic acid molecule or a polypeptide encoded thereby for identifying substances capable of interacting with said polypeptide.

33. The use of claim 32, wherein said substance(s) capable of interacting with said polypeptide is/are (an) antagonist(s) or (an) agonist(s).

- 86 -

34. A method of identifying a polypeptide or (a) substance(s) involved in cellular metabolism in an animal or capable of modifying homeostasis comprising the steps of:

- 5 (a) testing a collection of polypeptides or substances for interaction with a polypeptide of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, 10 KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or formin-binding protein 21 gene family or (a) fragment(s) thereof using a readout system; and
- (b) identifying polypeptides or substances which test positive for 15 interaction in step (a).

35. A method of identifying a polypeptide or (a) substance(s) involved in cellular metabolism in an animal or capable of modifying homeostasis comprising the steps of

- 20 (a) testing a collection of polypeptides or substances for interaction with the polypeptide identified by the method of claim 34; and
- (b) identifying polypeptides that test positive for interaction in step (a); and optionally
- 25 (c) repeating steps (a) and (b) with the polypeptides identified one or more times wherein the newly identified polypeptide replaces the previously identified polypeptide as a bait for the identification of a further interacting polypeptide.

30 36. The method of claim 34 or 35 further comprising the step of identifying the nucleic acid molecule(s) encoding the one or more interacting (poly)peptides.

- 87 -

37. A method of identifying a (poly)peptide involved in the regulation of body weight in a mammal comprising the steps of

- 5 (a) contacting a collection of (poly)peptides with a polypeptide of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or formin-binding protein 21 gene family or (a) fragment(s) thereof under conditions that allow binding of said (poly)peptides;
- 10 (b) removing (poly)peptides from said collection of (poly)peptides that did not bind in step (a); and
- 15 (c) identifying (poly)peptides that bind.

38. The method of claim 37 wherein said polypeptide of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or formin-binding protein 21 family is fixed to a solid support.

20

25 39. The method of claim 38 wherein said solid support is a gel filtration or an affinity chromatography material.

40. The method of any one of claims 37 and 39 wherein, prior to said identification in step (c), said binding (poly)peptides are released.

30

41. The method of claim 40 wherein said release is effected by elution.

- 88 -

42. The method of any one of claims 37 to 41 further comprising the step of identifying the nucleic acid molecule(s) encoding the one or more binding (poly)peptides.

5 43. A method of identifying a compound influencing the expression of a nucleic acid molecule of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940  
10 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or formin-binding protein 21 gene family comprising the steps of  
(a) contacting a host cell or a non-human host carrying an expression vector of claim 20 or the nucleic acid molecule identified by the method of claim 36 or 42 operatively linked  
15 to a readout system with a compound or a collection of compounds;  
(b) assaying whether said contacting results in a change of signal intensity provided by said readout system; and, optionally,  
(c) identifying a compound within said collection of compounds  
20 that induces a change of signal in step (b);  
wherein said change in signal intensity correlates with a change of expression of said nucleic acid molecule.

25 44. A method of identifying a compound influencing the activity of a polypeptide of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or  
30 formin-binding protein 21 family comprising the steps of  
(a) contacting a non-human host or a host cell carrying an expression vector of claim 20 operatively linked to a readout

- 89 -

system and/or carrying a (poly)peptide of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or formin-binding protein 21 family linked to a readout system with a compound or a collection of compounds;

- (b) assaying whether said contacting results in a change of signal intensity provided by said readout system; and, optionally
- (c) identifying a compound within said collection of compounds that induces a change of signal in step (b);

wherein said change in signal correlates with a change in activity of said (poly)peptide.

45. The method of claim 43 or 44 wherein said host cell is a eukaryotic host cell, particularly a mammalian host cell.

46. The method of claim 43 or 44 wherein said host cell is a unicellular organism, particularly a bacterium or a yeast.

47. The method of any one of claims 43 to 46 wherein said change in signal intensity is an increase or decrease in signal intensity.

48. A method of assessing the impact of the expression of one or more polypeptides of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or formin-binding protein 21 family in an animal comprising the steps of

- 90 -

- (a) overexpressing a nucleic acid molecule of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or formin-binding protein 21 gene family or a nucleic acid molecule of claim 36 or 42 in said animal; and
- (b) determining whether the weight of said animal has increased, decreased, whether metabolic changes are induced and/or whether the eating behaviour is modified.

49. A method of assessing the impact of the expression of one or more polypeptides of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or formin-binding protein 21 family in an animal comprising the steps of

- (a) underexpressing the nucleic acid molecule of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or formin-binding protein 21 gene family or a nucleic acid molecule of claim 36 or 42 in said animal; and
- (b) determining whether the weight of said animal has increased, decreased, whether metabolic changes are induced and/or whether the eating behaviour is modified.

- 91 -

50. A method of screening for an agent which modulates the interaction of a polypeptide of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or formin-binding protein 21 family with a binding target/agent, comprising the steps of

(a) incubating a mixture comprising

(aa) said polypeptide or a fragment thereof;

(ab) a binding target/agent of said (poly)peptide or fragment thereof; and

(ac) a candidate agent

under conditions whereby said (poly)peptide, or fragment thereof specifically binds to said binding target/agent at a reference affinity;

(b) detecting the binding affinity of said (poly)peptide, or fragment thereof to said binding target to determine an (candidate) agent-biased affinity; and

(c) determining a difference between (candidate) agent-biased affinity and the reference affinity.

51. A method of screening for an agent which modulates the activity of a polypeptide of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or formin-binding protein 21 family comprising the steps of

(a) incubating a mixture comprising

(aa) said polypeptide or a fragment thereof; and

- 92 -

- (ab) a candidate agent  
under conditions whereby said (poly)peptide, or fragment  
thereof has a reference activity,
- (b) detecting the activity of said (poly)peptide, or fragment  
thereof to determine an (candidate) agent-biased activity; and
- (c) determining a difference between (candidate) agent-biased  
activity and the reference activity.
52. A method of refining the compound identified by the method of any  
one of claims 43 to 47 or the agent identified by the method of  
claim 50 or 51 comprising
- (a) modeling said compound by peptidomimetics; and
- (b) chemically synthesizing the modeled compound.
53. A method of producing a composition comprising formulating the  
compound identified by the method of any one of claims 43-47 or  
the agent identified by the method of claim 50 or 51 or the  
compound refined by the method of claim 52 with a  
pharmaceutically acceptable carrier and/or diluent.
54. A method of producing a composition comprising the compound  
identified by the method of any one of claims 43 to 47 or the agent  
identified by the method of claim 50 or 51 comprising the steps of
- (a) modifying a compound identified by the method of any one of  
claims 43 to 47 or the agent of claim 50 or 51 as a head  
compound to achieve
- (i) modified site of action, spectrum of activity, organ  
specificity, and/or
- (ii) improved potency, and/or
- (iii) decreased toxicity (improved therapeutic index), and/or
- (iv) decreased side effects, and/or

- 93 -

- (v) modified onset of therapeutic action, duration of effect, and/or
- (vi) modified pharmacokinetic parameters (resorption, distribution, metabolism and excretion), and/or
- 5 (vii) modified physico-chemical parameters (solubility, hygroscopicity, color, taste, odor, stability, state), and/or
- (viii) improved general specificity, organ/tissue specificity, and/or
- 10 (ix) optimized application form and route, and
- (b) formulating the product of said modification with a pharmaceutically acceptable carrier.
- 15 55. The method of claim 53 or 54 wherein said composition is a pharmaceutical composition.
56. The method of claim 55, wherein said composition is a pharmaceutical composition for preventing, alleviating or treating
- 20 diseases and disorders related to body-weight regulation and thermogenesis, for example, but not limited to, metabolic diseases such as obesity, as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, and gallstones,
- 25 and disorders related to ROS defence, such as diabetes mellitus, neurodegenerative disorders, mitochondrial disorders and others.
57. A composition comprising
- 30 (a) an inhibitor or stimulator of the (poly)peptide of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or

- 94 -

epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or formin-binding protein 21 family or identified by the method of any one of claims 34, 35 or 37 to 41, 50 or 51 or refined by the method of claim 52;

- (b) an inhibitor of the expression of the gene identified by the method of claim 36 or 42; and/or
- (c) a compound identified by the method of claim 43 or 44.

58. The composition of claim 57 which is a pharmaceutical composition.

59. Use of

- (a) an inhibitor or stimulator of the (poly)peptide identified by the method of any one of claims 34, 35, 37 to 41 or 43 to 47, 50 or 51 or refined by the method of claim 52;
- (b) an inhibitor or stimulator of the expression of the gene identified by the method of claim 36 or 42; and/or
- (c) a compound identified by the method of claim 47;

for the preparation of a pharmaceutical composition for the treatment of obesity, as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis and gallstones and disorders related to ROS defence, such as diabetes mellitus, neurodegenerative disorders, mitochondrial disorders and others.

60. Use of an agent as identified by the method of claim 50 or 51 for the preparation of a pharmaceutical composition for the treatment, alleviation and/or prevention of obesity, as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia,

- 95 -

osteoarthritis and gallstones and disorders related to ROS defence, such as diabetes mellitus, neurodegenerative disorders, mitochondrial disorders and others.

- 5 61. Use of a nucleic acid molecule as depicted in SEQ ID NO:4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, and 38 or of (a) fragment(s) thereof for the preparation of a non-human animal which over- or underexpresses the gene product as encoded by said nucleic acid.
- 10 62. Kit comprising at least one of
- (a) a nucleic acid molecule of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or formin-binding protein 21 gene family;
  - (b) a vector of claim 20;
  - 20 (c) a host of claim 21;
  - (d) a polypeptide of the Optic atrophy 1 protein, cornichon-like, IGF-II mRNA-binding protein 3, neuralized-like, KIAA1094 protein, casein kinase (delta or epsilon), glutamate dehydrogenase, kraken homolog, sirtuin 1, escargot homolog, KIAA1585 protein, CG11940 homolog, dappled homolog, CG11753 homolog, KIAA0095 protein, and/or formin-binding protein 21 family;
  - 25 (e) a fusion protein of the polypeptide (d);
  - (f) an antibody or a fragment or derivative thereof or an antiserum, an aptamer or another receptor of claim 23; and
  - 30 (g) a hybridization probe, primer or anti-sense oligonucleotide of claim 24.

**FIGURE 1: Drosophila UCPy****FIGURE 1A: full length cDNA (SEQ ID NO:1)**

CGAGNAAGTGTACTATCTAAACACATTTCAAACAATCTTAACAAACAATTCCAAACATACAAATCCACTTACCACTTA  
CCGACCAAATACGAGTTTACAATGGACAAAGCTGAACGCGACTACTGGCATCTTCGATCCTTGGAAATCGAAGAGGAGC  
CGCGATTTCCGCCAACAAACGTCGCTGATCCACTAACCGCACGCAATCTGTTCCAGCTCTACGTCAACACCTTCATTGGA  
GCCAATCTGGCCGAGTCGTGTTTCCCATTTGGACCTGGCCAAAGACCCGGATGCAGGTAGATGGCGAGCAGGCCAAGAA  
GACGGGTAAAGCGATGCCAACTTTCCGTGCAACTCTTACCAACATGATCCGAGTGGAGGGATTCAAGTCGCTCTACGCCG  
GCTTCTCGGCAATGGTGACCCGAAACTTTATCTTCAACTCGTTACGTGTTGTTCTCTACGACGTTTTCCGGCgCCCTTTT  
CTCTACCAGAACGAACGGAACGAGGAAGTGCTCAAGATCTACATGGCGCTGGGATGCAGCTTCACCGCAGGCTGCATTGC  
CCaGGCACTGgCCAATCCcTTTGACATcGTCAAGGTGCGaATGCAGacGGAAGgaCgCCgCCGCCAGcTGGgcTATGATG  
TGCGGGTgAACAGCATGGTGCAAGCcTtCGTGgACATCTACCGCcGTGGCGGAcTGCCCAAGTATGTGgAAGGGTGTAGGg  
CCCAGCTGCATGCGTGCCCTGCTGATGACGACCGCGCATGTGGgCAGTTACgATATCAGTAAGCGCACCTTCAAGCGCcT  
GcTGGACTTGGAGGAAGGCCCTGCCACTGcGTTTcGTGTcTTCATGtGCGCCGGACTAACGGCATCCGTGCTCAGCACGC  
CGCGCaACGTGATCAAGTTCGCGGATGATGaaCCAGCCGGTGaACGAGAGCGGCAAGAATcTGTACTACAAGAATCCCTC  
GacTGCATTAGGAAGCTGGTCAGGGAGGAGGGTGTCTTCACGTGTATAAGGGCTCATGCCAcTTGGTTTCGCTGGG  
ACCGTTCTCAGTGCTCTTTTGGCTGTCCGTCGAGCAGCTGCGTCAGTGGaAAGGCCAGAGTGGATTTTAGGAGCAAACTA  
TCAATCTTACTATCGTATTTTGTATGTCTTTTAACACGCAATAAAAGGGTGCAAGTCAAACCATCTATTATACATATTA  
TAAATATAaCTTTAATCCCAAAAAAAAAAAAAAAAACTCGTGCCGAATTCGAT

**FIGURE 1B: open reading frame (SEQ ID NO:2)**

ATGGACAAAGCTGAACGCGACTACTGGCATCTTCGATCCTTGGAAATCGAAGAGGAGCCGCGATTTCCGCCAACAAACGT  
CGCTGATCCACTAACCGCACGCAATCTGTTCCAGCTCTACGTCAACACCTTCATTGGAGCCAATCTGGCCGAGTCGTGTG  
TTTTCCCATTTGGACGTGGCCAAAGACCCGGATGCAGGTAGATGGCGAGCAGGCCAAGAAGACGGGTAAAGCGATGCCAACT  
TTCCGTGCAACTCTTACCAACATGATCCGAGTGGAGGGATTCAAGTCGCTCTACGCCGGCTTCTCGGCAATGGTGACCCG  
AAACTTTATCTTCAACTCGTTACGTGTTGTTCTCTACGACGTTTTCCGGCgCCCTTTTCTCTACCAGAACGAACGGAACG  
AGGAAGTGTCAAGATCTACATGGCGCTGGGATGCAGCTTCACCGCAGGCTGCATGCCCCaGGCACTGgCCAATCCcTTT  
GACATcGTCAAGGTGCGaATGCAGacGGAAGgaCgCCgCCGCCAGcTGGgcTATGATGTGCGGGTgAACAGCATGGTGCA  
GGCcTtCGTGgACATCTACCGCcGTGGCGGAcTGCCCAAGTATGTGgAAGGGTGTAGGgCCCAGCTGCATGCGTGCTGCC  
TgATGACGACCGCGCATGTGGgCAGTTACgATATCAGTAAGCGCACCTTCAAGCGCcTgCTGGACTTGGAGGAAGCCCTG  
CCACTGcGTTTTcGTGTcTTCATGtGCGCCGGACTAACCGCATCCGTGCTCAGCACGCCGGCGaACGTGaTCAAGTCGCG  
GATGATGAaCCAGCCGGTGaACGAGAGCGGCAAGAATcTGTACTACAAGAATCCCTCGacTGCATTAGGAAGCTGGTCA  
GGGAGGAGGGTGTCTTCACGTGTATAAGGGCTCATGCCAcTTGGTTTCGCTGGGACCGTTCTCAGTGCTCTTTTGG  
CTGTCCGTGCGAGCAGCTGCGTCAGTGGaAAGGCCAGAGTGGATTTTAG

**FIGURE 1C: amino acid sequence encoding UCPy (SEQ ID NO:3)**

MDKAERDYWHLSLEIEEPRFPPTNVADPLTARNLFQLYVNTFIGANLAESCVFPLDVAKTRMQVDGEQAKKTGKAMPT  
FRATLTNMI RVEGFKSLYAGFSAMVTRNFI FNSLRVLYDVFRFPFLYQNERNEEVLKIYMALGCSFTAGCIAQALANPF  
DIVKVRMQTEGRRRQLGYDVRVNSMVQAFVDIYRRGGLPSMWKGVGPSCMRACLMTTGDVGSYDISKRTFKRLLDLEEGL  
PLRFVSSMCAGLTASVLSTPANVIKSRMMNQPVNESGKNLYYKNSLDCIRKLVREEGVLTLYKGLMPTWFRLLGPF SVLFW  
LSVEQLRQWKQSGF

**FIGURE 2: HUMAN HOMOLOG OF CG8479****FIGURE 2A: Homology to human gene**

ref|XP\_039926.2| (XM\_039926) KIAA0567 protein

/gariid=G2CZX6H97482VL /chrom=3 /contig=NT\_005571.3 /start=532481 /end=647399  
 /strand=minus OPA1: optic atrophy 1 (autosomal dominant) (optic atrophy 1  
 gene;KIAA0567) Length = 11424

Score = 45.7 bits (106), Expect = 7e-04  
 Identities = 61/99 (61%), Positives = 77/99 (77%), Gaps = 65/99 (65%)  
 Frame = +3

Query: 313 DGSVDA-RSNVTDVMCD----GRRTV---TKVDAAD-----DRRK 344  
 DGSVDA RS VTD++ GRRT+ TKVD A+ ++ +  
 Sbjct: 5277 DGSVDAERSIVTDLVSQMDPHGRRTIFVLTKVDLAENVASPSRGKACRGINGSMEKGR 5456

Query: 345 SGK-----MKA-GYYAVVTGRGRKDDSD-----DARYDKNSKHRRGVM---- 380  
 S K MKA GY+AVVTG+G +S + + +NSK + M  
 Sbjct: 5457 SEK\*IQQIIEGKLFPMKALGYFAVVTGKGNSSSEIEAIREYEEFFQNSKLLKTSMLKAH 5636

Query: 381 HVTSRN--SAVSD-RWKMVR-----TADA-KATR----NTWKNN 411  
 VT+RN AVSD WKMVR AD+ KATR WKNN  
 Sbjct: 5637 QVTTRNLSLAVSDCFWKMVRESVEQQADSFKATRFNLETEWKNN 5768

**FIGURE 2B: Predicted coding sequence for the human homolog protein (1975 bp)  
(SEQ ID NO:4)**

&gt;DTT02151012

ATGTGGCGACTACGTCGGGGCCGCTGTGGCCCTGTGAGGCTGCCAGTCTTTAGTGAAACACAGCTCTGGAA  
 TAAAAGGAAGTTTACCACTACAAAACTACATCTGGTTTCACGAAGCATTATCATTACATCATCCTAC  
 CTTAAAGCTTCAAGGACCCCAATTAAGGACATCCTTTTCAGCAGTTCTCTCTCTGACAAACCTTCCTTTA  
 CGTAAACTGAAATCTCTCCAATTAATATGGCTACCAGCCTCGCAGGAATTTTGGCCAGCAAGATTAG  
 CTACGAGACTCTTAAACTTCGCTATCTCATACTAGGATCGGCTGTTGGGGGTGGCTACACAGCCAAAAA  
 GACTTTTGATCAGTGGAAGATATGATACCGGACCTTAGTGAATATAAATGGATTGTGCCTGACATTGTG  
 TGGGAAATTGATGAGTATATCGATTTTTGAGAAAATTAGAAAAGCCCTTCTTAGTTTCAGAAAGACCTTGTA  
 AGTTAGCACCAGACTTTGACAAGATTGTTGAAAGCCTTAGCTTATTGAAGGACTTTTACCTCAGGTTT  
 TCCGGAAGAAACGGCGTTTAGAGCAACAGATCGTGGATCTGAAAGTGACAAGCATTTAGAAAGGGTCTG  
 CTGCTGAGCTCATTCTCTTACAACAACAAATTCAGAGCATGAAGAGGAAGCGCGCAGAGCCGCTGGCC  
 AATATAGCACGAGCTATGCCCCAACAGAAGCGCAAGGTGTCAGACAAAGAGAAAATTGACCAACTTCAGGA  
 AGAACTTCTGCACACTCAGTTGAAGTATCAGAGAATCTTGAACGATTAGAAAAGGAGAACAAAGAAATG  
 AGAAAATTAGTATGTCAGAAAGATGACAAAGGCATTCATCATAGAAAGCTTAAGAAATCTTTGATTGACA  
 TGATTCTGAAGTTCTTGATGTTCTCTCTGATTATGATGCCAGTTATAATACGCAAGATCATCTGCCACG  
 GGTGTTGTGTTGGAGATCAGAGTGCTGGAAGACTAGTGTGTTGGAAATGATTGCCCAAGCTCGAATA  
 TTCCCAAGAGGATCTGGGGAGATGATGACACGTTCTCCAGTTAAGGTGACTCTGAGTGAAGGTCTCACC  
 ATGTGGCCCTATTTAAAGATAGTTCTCGGGAGTTTGATCTTACCAAGAAGAAGATCTGTCAGCATTAG  
 ACATGAAATAGAACTTCGAATGAGGAAAAATGTGAAAGAAGGCTGTACCGTTAGCCCTGAGACCATATCC  
 TTAATGTAAGGCTTACAGACTACAGAGGATGGTGTGTTGACTTACCAGGTGTGATTAATACTGTGA  
 CATCAGGTCATGCTCCTGACACAAAGGAACTATTTTCAGTATCAGCAAAGCTTACATGTCAGAAATCCTAA  
 TGCCATCATACTGTGTATTCAAGATGGATCTGTGGATGCTGAACGCAGTATTGTTACAGACTTGGTCAGT  
 CAAATGGACCTCATGGAAGGAGAACCATATTCGTTTTGACCAAAGTAGACCTGGCAGAGAAAAATGTAG  
 CCAGTCCAAGCAGGATTCAGCAGATAATTGAAGGAAAGCTCTTCCCAATGAAAGCTTAGGTTATTTTGC  
 TGTGTGAACAGGAAAGGAAAGCCTCTGAAGCAATTGAAGCTATAAGAGAAATATGAAGAAGAGTTTTT  
 CAGAATTCAAAGCTCCTAAAGACAAGCATGCTAAAGGCACACCAAGTGACTACAAGAAATTTAAGCCTTG  
 CAGTATCAGACTGCTTTTGGAAAATGGTACGAGAGTCTGTTGAACAACAGGCTGATAGTTTCAAAGCAAC  
 ACGTTTTAACCTTGAAACTGAATGGAAGAATAACTATCCTCGCCTGCGGGAACCTTGACCGGTAATATTT  
 GGATACTCCTAAAAATGAAATCCTTGATGAAGTTATCAGTCTGAGCCAGGTTACACCAAAACATTGGGAG  
 GAAATCCTTCAATAA

**FIGURE 2C: Predicted amino acid sequence for the human homolog protein (658 aa)  
(SEQ ID NO:5)**

>DTP02151021

MWRLRRAAVACEVCQSLVKHSSGIKGSPLQKLHLVSRSIYHSHHPTLKLQRPQLRTSFQQFSSLTNLPL  
RKLKFSPIKYGYQPRRNFWPARLATRLKLRLYLILGSAVGGGYTAKKTFDQWKDMIPDLSEYKWIVPDIV  
WEIDEYIDFEKIRKALPSSDLVKLAPDFDKIVESLSLLKDDFTSGSPEETAFRATDRGSESDKHFRKGL  
LGELILLQQQIQEHEEEARRAAGQYSTSYAQQRKVSDEKIDQLQEELLHTQLKYQRILERLEKENKEL  
RKLVLQKDDKGIHHRKLKSLIDMYSEVL DVLS DYDASYNTQDHLPRVVVVGDSAGKTSVLEMIAQARI  
FPRGSGEMMTRSPVKVTLSEGP HHVALFKDSSREFDLTKEEDLAALRHEIELMRKNVKEGCTVSPETIS  
LNVKGPGLQRMVLVDLPGVINTVTSGMAPDTKETIFSISKAYMQNPNAIILCIQDGSVDAERSIVTDLVS  
QMDPHGRRTIFVLTKVDLAEKNVASPSRIQQIIEGKLFPMKALGYFAVVTGKGNSSSEIEAIREYEEFF  
QNSKLLKTSMLKAHQVTTRNL SLAVSDCFWKMVRESVEQQADSFKATRFNLETEWKNNYPRLRELD RVIF  
GYXXKNEILDEVISLSQVTPKHWEIILQ\*

**FIGURE 3: Multiple Sequence Alignment (ClustalW 1.83)**

```

OPA1-5 Hs      MWRLRRAAVACEVCQSLVKHSSGIKGSPLQLKLHLVSRSIYHSHHPTLKLQRPQLRTSFQ
XP_148016 Mm   MWRAGRAAVACEVCQSLVKHSSGIQRNVPLQKLHLVSRSIYRSHHPALKLQRPQLRTFFQ
CG8479 Dm      MLRIYQNTYRRTARKAVVYSTK--VACCNHSTLCGITSHPRRAQDSGSSSSNGRHRGHEE

OPA1-5 Hs      QFSSLTNLPLRKLKFSPIKYGYQPRRNFWPARLATRLKLRLYLILGSAVGGGYTAKKTFD
XP_148016 Mm   QFSSLTHLSLHKLKLSPIKYGYQPRRNFWPARLAARLLKLRYIILGSAVGGGYTAKKTFD
CG8479 Dm      FLLAGNPARGWQMP--PPSRGYG---MLVVRILRGALKLRYIVLGAIGGGVSLSKKYE

OPA1-5 Hs      QWKDMIPDLSEYKWIVPDIWWEIDEYIDFEKIRKALPNSEDLVKLAPDFDKIVES-LSL
XP_148016 Mm   EWKDMIPDLSDYKWIVPDIWWEIDEYIDLEKIRKALPSEDLASLAPDLDKITES-LSL
CG8479 Dm      EWKDGLPNFKWLEDAMPQGERWSQFSRNLEIVGSLVKNA---IEVDPKCLKQLGEDKLSW

OPA1-5 Hs      KDFFTSGHKLVEVIGASDLLLLGSPETAFR--ATDRGSESDKHFRKVSDE-KIDQL
XP_148016 Mm   KDFFTAGPKLVSEVLEVSEALLLLGSPETAFR--ATDHGSESDKHFRKVSDE-KIDQL
CG8479 Dm      RNWFDSRLDDAIEAADYQG-VQIVETKDDLKAKTTVAALGITSDESRRKYEKLQSQVETL

OPA1-5 Hs      QEELLHTQLKYQRILERLEKENKELRK--LVLQKDDKGIHHRKLKSLIDMYSEVLVLVS
XP_148016 Mm   QEELLHTQLKYQRILERLEKENKELRK--LVLQKDDKGIHHRKLKSLIDMYSEVLVLVS
CG8479 Dm      QTEIMNVQIKYQKELEKMEKENRELQQYLILKTN-KKTTAKKIKKSLIDMYSEVLDELS

OPA1-5 Hs      DYDASYNTQDHLPRVVVVDQSGAKTSVLEMIQAARIFPRGSGEMMTRSPVKVTLSEGP
XP_148016 Mm   DYDASYNTQDHLPRVVVVDQSGAKTSVLEMIQAARIFPRGSGEMMTRSPVKVTLSEGP
CG8479 Dm      GYDTGYTMADHLPRVVVVDQSSGKTSVLESIAKARIFPRGSGEMMTRAPVKVTLAEGPY

OPA1-5 Hs      HVALFKDSSREFDLTKEEDLAALRHEIELMRMKNVKEGCTVSPETISLNVKGPGLQRMVL
XP_148016 Mm   HVALFKDSSREFDLTKEEDLAALRHEIELMRMKNVKEGCTVSPETISLNVKGPGLQRMVL
CG8479 Dm      HVAQFRDSDREYDLTKEEDLQDLRRDVEFRMKASVRGGKTVSNEVIAMTVKGPGLQRMVL

OPA1-5 Hs      VDLPGVINTVTSGMAPDTKETIFSISKAYMQNPNAIILCIQDGSVDAERSIVTDLVSQMD
XP_148016 Mm   VDLPGVINTVTSGMAPDTKETIFSISKAYMQNPNAIILCIQDGSVDAERSIVTDLVSQMD
CG8479 Dm      VDLPGIISTMTVDMASTKDSIHQMTKHYMSNPNAIILCIQDGSVDAERSNVTDLVMQCD

OPA1-5 Hs      PHGRRTIFVLTKVDLAEKNVASPSRIQQIIEGKLFPMKALGYFAVVTGKGNSSSEIEAIR
XP_148016 Mm   PHGRRTIFVLTKVDLAEKNVASPSRIQQIIEGKLFPMKALGYFAVVTGKGNSSSEIEAIR
CG8479 Dm      PLGRRTIFVLTKVDLAE--LADPDRIKILSGKLFPMKALGYAVVTGRGRKDDSIDAIR

OPA1-5 Hs      EYEEFFQNSKLLKT-SMLKAHQV'TTRNLSLAVSDCFWKMVRESVEQQADSFKATRFNLE
XP_148016 Mm   EYEEFFQNSKLLKT-SMLKAHQV'TTRNLSLAVSDCFWKMVRESVEQQADSFKATRFNLE
CG8479 Dm      QYEEFFQNSKLFHRRGVIMPHQVTSRNLSLAVSDRFWKMVRETIEQQADAFKATRFNLE

OPA1-5 Hs      TEWKNNYPRLRELDRLNLEFEKAKNEILDEVISLSQVTPKHWEELQQLSLWERVSTHVIEN
XP_148016 Mm   TEWKNNYPRLRELDRLNLEFEKAKNEILDEVISLSQVTPKHWEELQQLSLWERVSTHVIEN
CG8479 Dm      TEWKNNYPRLRESGRDELFDKAKGEILDEVVTLSQLSARKWDDALSTKLWEKLSNYVFES

OPA1-5 Hs      IYLPAAQTMNSGTFNTTVDIKLKQWTDKQLPNKAVEVAVETLQEEFSRFMTEPK-GKEHD
XP_148016 Mm   IYLPAAQTMNSGTFNTTVDIKLKQWTDKQLPNKAVEVAVETLQEEFSRFMTEPK-GKEHD
CG8479 Dm      IYLPAAQSGSQNSFNTMVDIKLRQWAEQALPAKSVEAGWEALQEEFISLMERSKKAQDHD

OPA1-5 Hs      DIFDKLKEAVKEESIKRHKWNDFAEDSLRLVIQHNALEDRSISDKQQWDAAIYFMEEALQA
XP_148016 Mm   DIFDKLKEAVKEESIKRHKWNDFAEDSLRLVIQHNALEDRSISDKQQWDAAIYFMEEALQA
CG8479 Dm      GIFDQLKSAVVDEAIRHSHWEDKAIDMLRLVIQLNTLEDRFVHDKQEWDSAVKFLESSVNA

OPA1-5 Hs      RLKDTENAIENMVGPDWKKRWLYWKNRTQEQCVHNETKNELEKMLKCNEEHPAYLASDEI
XP_148016 Mm   RLKDTENAIENMIGPDWKKRWLYWKNRTQEQCVHNETKNELEKMLKVNDEHPAYLASDEI
CG8479 Dm      KLVQTEETLAQMFGPGQMRRI THWQYLTQDQKRRSVKNELDKILKNDTKHLPTLTHDEL

OPA1-5 Hs      TTVRKNLESRGVEVDPSLIKDTWHQVYRRHFLKTALNHCNLCRRGFYYYQRHFVDSLEEC
XP_148016 Mm   TTVRKNLESRGVEVDPSLIKDTWHQVYRRHFLKTALNHCNLCRRGFYYYQRHFIDSELEC
CG8479 Dm      TTVRKNLQRDNVDVDTDYIRQTFWPVYRKHFLQQALQRAKDCRKAYLYTQQGAECEISC

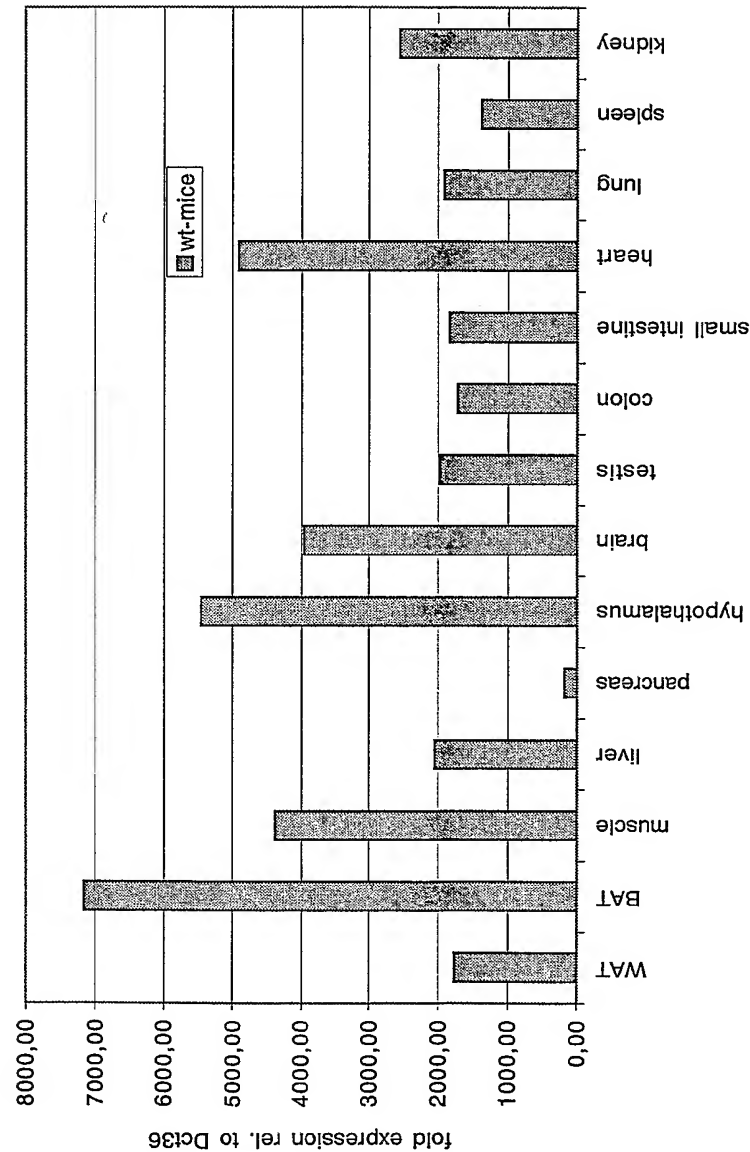
```

OPA1-5 Hs NDVVLFWRIQRM LAITANTLRQQLTNTTEVRRLEKNVKEVLEDF AEDGEKKIKLLTGKRVQ  
XP\_148016 Mm NDVVLFWRIQRM LAITANTLRQQLTNTTEVRRLEKNVKEVLEDF AEDGEKKVKLLTGKRVQ  
CG8479 Dm SDVVLFWRIQQVIKITGNALRQQVINREARRLDKEIKAVLDEFSDDEEKKGYLLTGKRVL

OPA1-5 Hs LAEDLKKVREIQEKLD AFIEALHQEK  
XP\_148016 Mm LAEDLKKVREIQEKLD AFIEALHQEK  
CG8479 Dm LAEELIKVRQIQEKLEEFINSLNQEK

FIGURE 4. Expression of OPA1 in mammalian tissues

FIGURE 4A. Real-time PCR analysis of OPA1 expression in wildtype mouse tissues



**FIGURE 5: HUMAN HOMOLOG OF CG5855 (cornichon)****FIGURE 5A: BLASTN SEARCH RESULT****Homology to human gene ref|NP\_005767.1|**

/protein=DTF09557033.1 /gene=DTG09557004.1 /locus=DTL09557002.1  
 /garid=G2Q19PL\_9GKV875 /chrom=14 /contig=NT\_010140.3  
 /start=12537939 /end=12551452 /strand=plus Similar to:  
 gi|5031639|ref|NP\_005767.1| cornichon-like [Homo sapiens] Length = 435

Score = 202 bits (508), Expect = 2e-52  
 Identities = 91/144 (63%), Positives = 111/144 (76%)  
 Frame = +1

Query: 1 MAFNFTAFTYIVALIGDAFLIFFAIFHVIAFDELKTDYKNPIDQCNSLNPLVLPEYXXXX 60  
 MAF F AF Y++AL+ A LIFFAI+H+IAFDELKTDYKNPIDQCN+LNPLVLPEY  
 Sbjct: 1 MAFTFAAFCYMLALLLTAALIFFAIWHIIAFDELKTDYKNPIDQCNTLNPLVLPEYLIHA 180

Query: 61 XXXXXXXXXCGEWFSLCINIPLIAYHIWRYKNRPVMSGPGLYDPTTVLKTDTLYRNMREGW 120  
 EW +L +N+PL+AYHIWRY +RPVMSGPGLYDPT++ D L +EGW  
 Sbjct: 181 FFCVMFLCAAELWLTGLNMPLLAYHIWRYMSRPVMSGPGLYDPTTIMNADILAYCQKEGW 360

Query: 121 IKLAVYLISFFYYIYGMVYSLIST 144  
 KLA YL++FFYY+YGM+Y L+S+  
 Sbjct: 361 CKLAFYLLAFFYYLYGMIYVLVSS 432

**FIGURE 5B: Predicted nucleotide sequence encoding the human homolog (435 bp)  
(SEQ ID NO:6)**

>DTF09557024

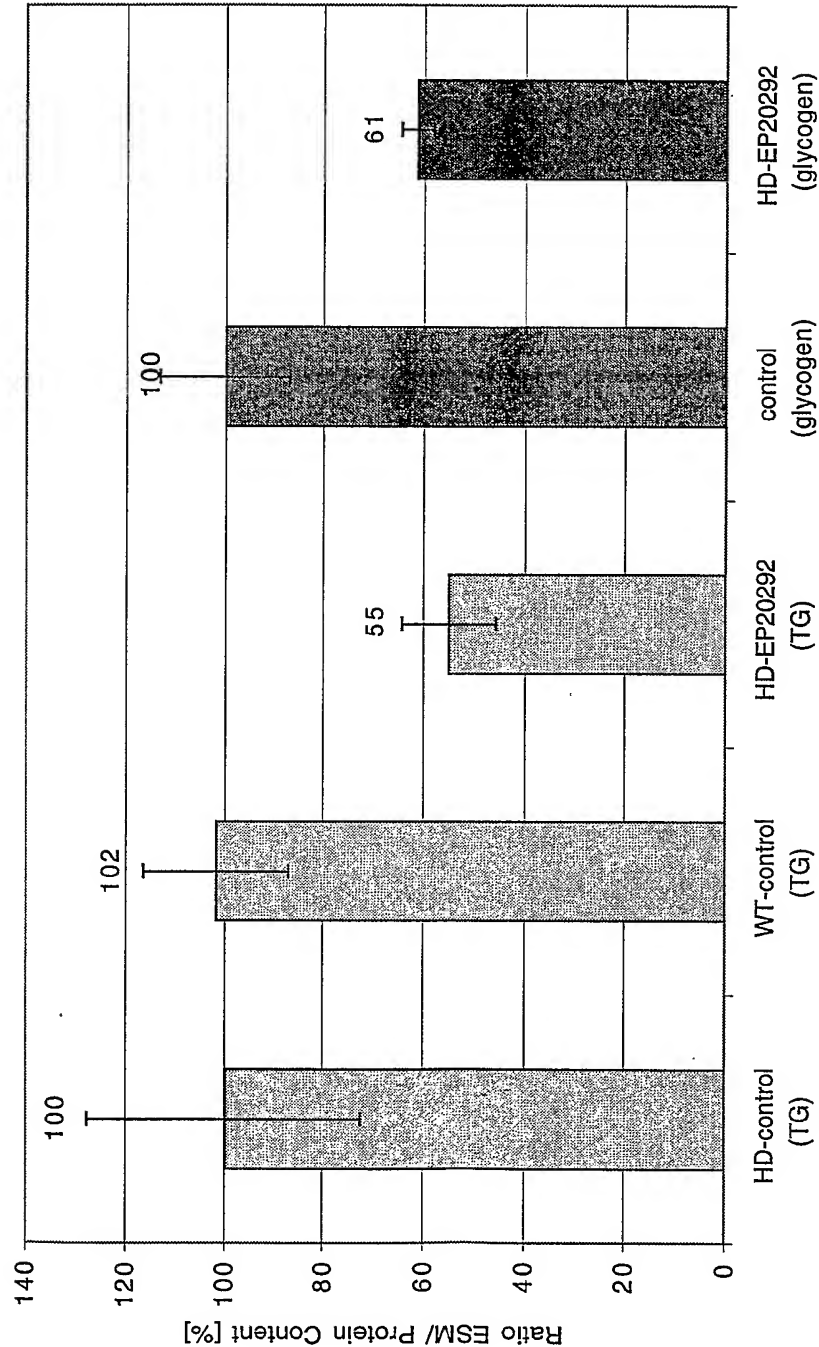
ATGGCGTTCACGTTTCGCGGCCTTCTGCTACATGCTGGCGCTGCTGCTCACTGCCGCGCTCATCTTCTTCG  
 CCATTTGGCACATTATAGCATTTGATGAGCTGAAGACTGATTACAAGAATCCTATAGACCAGTGTAATAC  
 CCTGAATCCCCCTTGTA TCTCCAGAGTACCTCATCCACGCTTTCTTCTGTGTCATGTTTCTTTGTGCAGCA  
 GAGTGGCTTACACTGGGTCTCAATATGCCCCCTCTTGGCATATCATATTTGGAGGTATATGAGTAGACCAG  
 TGATGAGTGGCCCAGGACTCTATGACCCTACAACCATCATGAATGCAGATATTCTAGCATATTGTCAGAA  
 GGAAGGATGGTGCAAATTAGCTTTTATCTTCTAGCATTTTTTTTACTACCTATATGGCATGATCTATGTT  
 TTGGTGAGCTCTTAG

**FIGURE 5C: Predicted amino acid sequence of the human homolog Protein (144 aa)  
(SEQ ID NO:7)**

>DTP09557033

MAFTFAAFCYMLALLLTAALIFFAIWHIIAFDELKTDYKNPIDQCNTLNPLVLPEYLIHAFFCVMFLCAA  
 EWLTLGLNMPLLAYHIWRYMSRPVMSGPGLYDPTTIMNADILAYCQKEGWCKLAFYLLAFFYYLYGMIYV  
 LVSS\*

FIGURE 6: Energy storage metabolite (ESM) content of a cornichon (Gadfly Accession Number CG5855) Mutant



**FIGURE 7. HUMAN HOMOLOG OF CG1691 (Imp)****FIGURE 7A. BLASTN SEARCH RESULTS****Homology to human gene ref|NP\_006537.1|**

Similar to: gi|12733121|ref|XP\_004780.2| IGF-II mRNA-binding protein 3  
[Homo sapiens]  
Length = 1405

Score = 289 bits (731), Expect = 8e-78  
Identities = 159/347 (45%), Positives = 222/347 (63%), Gaps = 3/347 (0%)  
Frame = +1

Query: 139 PGMPGPGRQADFPLRILVQSEMVGAIIGRQGSTIRTITQQSRARVDVHRKENVGSLEKSI 198  
PG + D PLR+LV ++ VGAIIG++G+TIR IT+Q++++DVHRKEN G+ EKSI  
Sbjct: 82 PGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAEKSI 261

Query: 199 TIYGNPENCTNACKRILEVMQQEAISTNKGELSPECEICLKILAHNNLIGRIIGKSGNT 258  
TI PE + ACK ILE+M +EA E EI LKILAHNN +GR+IGK G  
Sbjct: 262 TILSTPEGTSACKSILEIMHKEAQDIKFTE-----EIPKLKILAHNNFVGRLLIGKEGRN 423

Query: 259 IKRIMQDQDITKITVSSINDINSFNLERIITVKGLIENMSRAENQISTKLRQSYENDLQAM 318  
+K+I QDQDITKIT+S + ++ +N ER ITVKG +E ++AE +I K+R+SYEND+ +M  
Sbjct: 424 LKKIEQDQDITKITISPLQELTYNPERTITVKGNVETCAKAEIEIMKKIRESYENDIASM 603

Query: 319 APQSLMFPGLHPMAM-MSTPGNGMVFNTPMPFSCQSFAMSKTPASVVPV--FPNDLQE 375  
Q+ + PGL+ A+ + P +GM TS P P+++ PP F E  
Sbjct: 604 NLQAHLPGLNLNALGLFPPTSGMPPTSGP-----PSAMTPYPQFEQSETE 747

Query: 376 TTYLYIPNNAVGAIIIGTRGSHIRSIMRFSNASLKIAPLDADKPLDQQTERKVTIVGTPEG 435  
T +L+IP +VGAIIG +G HI+ + RF+ AS+KIAP +A R V I G PE  
Sbjct: 748 TVHLFIPALSVGAIIIGKQGHIKQLSRFAGASIKIAPAEA----PDAKVRMVIITGPPEA 915

Query: 436 QWKAQYMI FEKMREEGFMCGTDDVRLTVELLVASSQVGRIGKGGQNVRE 485  
Q+KAQ I+ K++EE F+ ++V+L + V S GR+IGKGG+ E  
Sbjct: 916 QFKAQGRIYGIKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTASE 1065

**FIGURE 7B. Predicted coding sequence for the human homolog (564 bp) (SEQ ID NO:8)**

>DTT00108009

ATGAAGAAGCTGCGTGAGACCTTTGAAAATGATATGTTGGCTGTTAATACGCACTCCGGATACTTCTCCA  
GCTTGTACCCCCATCACCAGGTTGGCCCGTTCCCGCATCATCACTCTTATCCAGAGCAGGAGGTTGTGAA  
TCTCTTCATCCCAACCCAGGCTGTGGGCGCCATTATCAGGAAGAGGGGAGCACACATCAAACAGCTGGCG  
AGATTCGCCACAGCCTCCATCAAGATCGCCCTTGCGGAAGGCCAGACGTCACGAAAGGATGGCCATCA  
TCACCCGGCCACCGGAAGCCAGTTCAAGGCCAGGGACGGATCTTTGGGAAACTGAAAGAAGAAAACCTT  
CTTTAACCCCAAAGAAGAAGTGAAGCTGGAAGCCGTTATCAGAGTGCCCTCTTCCACAGCTGGCCGGGTG  
ATTGGCAAAGGTGTCAATACCTTGAATGAAGTGCAGAACTTAACCAGTGCAGAAAGTCATCGTGCCTCGTG  
ACCAAAGGCCAGATGAAAATGAGGAAGTGATCGTCAGAAATTATTGGACACTTCTTTGCTACCCCAACTGC  
ATAG

10/48

**FIGURE 7C. Predicted amino acid sequence for sequence for the human homolog Protein (187 aa) (SEQ ID NO:9)**

>DTP00108018

MKKLRETFENDMLAVNTHSGYFSSLYPHHQVGPFPHHHSYPEQEVVNLFIPTQAVGAIIRKGAHIKQLA  
RFATASIKIALAEGPDVNERMAITRPPEAQFKAQGRIFGKLKEENFFNPKEEVKLEARIRVPSSTAGRV  
IGKGVNTLNELQNLTSAEVIVPRDQRPDENEVIVRIIGHFFATPTA\*

11/48

**FIGURE 8. The human homolog of *neuralized* (GadFly Acc. No. CG11988)****FIGURE 8A. tBLASTN search result for *neuralized***

**Homology to human *neuralized* (*Drosophila*)-like gene ref NM\_004210; protein ref NP\_004201.1**

Length = 1674

Score = 196 bits (493), Expect = 8e-50  
Identities = 146/501 (29%), Positives = 235/501 (46%), Gaps = 9/501 (1%)  
Frame = +1

Query: 106 PLQFHS-VHGDNIRISRDTLARRFESFCRAITFSARPVRINERICVKFAEISNNWNGGI 164  
PL FH G I + +R SFC AITFS RPV I E++ +K + W+G +  
Sbjct: 133 PLLFHPHTKGSQILMDLSHKAVKRQASFCNAITFSNRPVLIYEQVRLKITKKQCCWSGAL 312

Query: 165 RFGFTSNDPVTLE-GTLPKYACPDLTNRPGFWAKALHEQYCEKDNILYVYVNGAGDVIY 223  
R GFTS DP + +LPKYACPD + GFWAKAL E++ + NI+ ++V+ G V +  
Sbjct: 313 RLGFTSKDPSRIHPDSLPHYACPDLVSGFWAKALPEEFANEGNIIAFWVDKKGRVFHR 492

Query: 224 INNEEKGVILTIGIDTRSLWTVIDIYGNCTGIEFLDSRIYMYQQQPAAIXXXXXXXXXXX 283  
IN+ + +G+ T LW ++D+YG G++ LDS + + P +  
Sbjct: 493 INDSAVMLFFSGVRTADPLWALVDVYGLTRGVQLLDSELVL----PDCL----- 627

Query: 284 XXXXXXXXXXXXXXXXXXXXQSRSLPGHTAAIEHDLERHVMPSLQSLHLAGNGGSVAS-- 341  
+S +L + E D + + SL L++ G G A+  
Sbjct: 628 -----RPRSFTALRRPSLRREAD-DARLSVSLCDLNVPGADGDEAAPA 753

Query: 342 ----VEQAAIAHDLANGLPPLRYNANGRLIPVPFHNTK-GRNVRLSQDRFVASRTESDFC 396  
+ Q ++ + LP +G L FH + G +VR+ ++ VA  
Sbjct: 754 AGCPIPQNSLNSQHSRALPA---QLDGD---RFHALRAGAHVRILDEQTVARVEHGRDE 915

Query: 397 QGVVFTARPIRIGELIVQLKTEQMYVGALALGLTSCNPAMLQPNLNDSDFLDRPE 456  
+ VFT+RP+R+ E + V+V ++ GAL+ G+T+C+P L+P DLP + L+DR E  
Sbjct: 916 RALVFTSRPVRVAETIFVKVTRSGGARPGALSFGVTTCDPGTLRPADLPFSPEALVDRKE 1095

Query: 457 YWVVSKDIAAAPQRGDEIAFFVAPNGEVSISKNNGPVVMHVDQSLQLWAFLDVYGSTQ 516  
+W V + + GD + V +GE+ +S N A + + VD S LW ++G+  
Sbjct: 1096FWAVCR-VPGLPHSGDILGLVFNADGELHLSHNGAAAGMQLCVDASQPLWMLFGLHGTIT 1272

Query: 517 SLRMFRQQLPNMVAYPSQPQVNVNXXXXXXXXXXXXXXXXXRLPMTESMSSLNAGATAKLLHH 576  
+R+ + PS LP + + + + A L  
Sbjct: 1273QIRILGSTILAERGIPS-----LPCSPASTPTSPSALGSRLSD 1386

Query: 577 PSQLSVAQSTSTLASAGGVNGSRMISMPN 606  
P LS S +SAGG + +S+P +  
Sbjct: 1387P-LLSTCSSGPLGSSAGGTAPNSPVSLPES 1473

Score = 78.4 bits (190), Expect = 3e-14  
Identities = 41/115 (35%), Positives = 59/115 (50%)  
Frame = +1

Query: 638 LAARPTATVTSSGVLACSSGTLISTSSQYIEQPIANSTNNAANKWXXXXXXXXXXXXX 697  
L P +T TS L G+ S L+ST SS + + N+  
Sbjct: 1324 LPCSPASTPTSPSAL-GSRLSDPLLSTCSSGPLGSSAGGTAPNSPVSLPESVTPGLGQW 1500

Query: 698 XAECTICYENPIDSVLYMCGHMCMDYDCAIEQWRGVGGGQCPLCRAVIRDVIRTY 752  
ECTICYE+ +D+V+Y CGHMC+CY C + + + CP+CR I+D+I+TY  
Sbjct: 1501 SDECTICYEHAVDTVIYTCGHMCLCYACGLRLKKAL-HACCPICRRPIKDIITY 1662

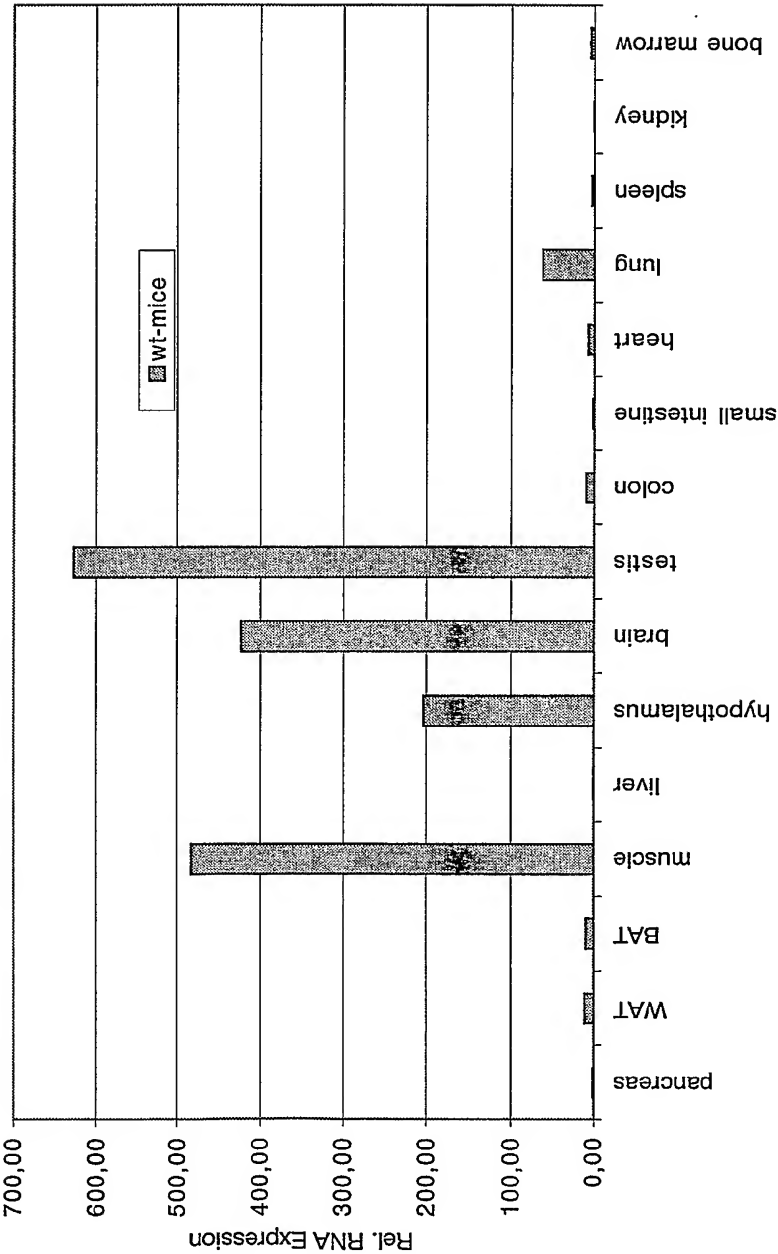
**FIGURE 8B. Predicted coding sequence for the human homolog of CG11988 (1674 base pairs); (SEQ ID NO:10)**

ATGGGGGGACAGATCACCCGGAGCACTCTCCACGACTCTATCGGGGGCCCCCTTCCCCGTCACCTTCACCC  
GATGCCACCACAAGCAGAAGCACTGTCCGGCAGTGCTGCCAGCGGGGGGCTCCCAGCCACGCCGCTGCT  
CTTCCACCCGCACACCAAGGGCTCCCAGATCCTCATGGACCTCAGCCACAAGGCTGTCAAGAGGCAGGCC  
AGCTTCTGCAACGCCATCACCTTCAGCAACCGCCCGGTCTCATCTACGAGCAAGTCAGGCTGAAGATCA  
CCAAGAAGCAGTGCTGCTGGAGCGGGGGCCCTGCGGCTGGGCTTCACCAGCAAGGACCCGCTCCCGCATCCA  
CCCTGACTCGCTGCCAAGTACGCCTGCCCCGACCTGGTGTCCCAGAGTGGCTTCTGGGCCAAGGCGCTG  
CCTGAGGAGTTTGCCAATGAGGGCAACATCATCGCATTCTGGGTGGACAAGAAGGGCCGTGTCTTCCACC  
GCATCAACGACTCGGCTGTTATGCTGTTCTTCAGCGGGGTCCGCACGGCCGACCCGCTCTGGGGCCCTGGT  
GGACGTCTACGGCCTCACGCGGGGCGTCCAGCTGCTTGATAGCGAGCTGGTGTCTCCGGACTGTCTGCGG  
CCGCGCTCCTTACCGCCCTGCGGCGGGCCGTGCTGCGGCGCGAGGCGGACGACGCGCGCCTCTCGGTGA  
GCCTATGCGACCTCAACGTGCCGGGCGCGGACGGCGACGAGGCCGCGCGCGGCGCGCGGCTGCCCCATCCC  
GCAGAACTCACTCAACTCGCAGCACAGCCGCGCGCTGCCGGCGCAGCTCGACGGCGACCTGCGTTTCCAC  
GCCCTGCGCGCCGGCGCGCACGTCCGCATCCTCGACGAGCAGACGGTGGCGCGCGTGGAGCACGGGCGCG  
ACGAGCGCGCGCTCGTCTTACCAGCCGGCCCGTGCCTGGCCGAGACCATCTTCGTCAAGGTCACGCG  
CTCGGGTGGCGCGCGGCCCGCGCGCTGTCTGCTCGGCGTACCACGTGCGACCCCGGCACGCTGCGGCGG  
GCCGACCTGCCTTTCAGCCCTGAGGCCCTGGTGGACCGCAAGGAATTCTGGGCCGTGTGCCGCGTGCCCG  
GGCCCCCTGCACAGCGGGCAGATCCTGGGCCTGGTGGTCAACGCCGACGGCGAGCTGCACCTCAGCCACAA  
TGGCGCGGCCCGCGGCATGCAGCTGTGCGTGGACGCCTCGCAGCCGCTTTGGATGCTCTTCGGCCTGCAC  
GGGACCATCACGCAGATCCGCATCCTCGGCTCCACTATCCTGGCCGAGCGGGGTATCCCATCACTCCCCT  
GCTCCCCCTGCCTCCACGCCAACCTCGCCAGTGCCCTGGGCAGCCGCTGTCTGACCCCTTGCTCAGCAC  
GTGCAGCTCTGGCCCTCTGGGTAGCTCTGCTGGTGGGACAGCCCCAATTGCCAGTGAGCCTGCCCGAG  
TCGCCAGTGACCCAGGTCTGGGCCAGTGGAGCGATGAGTGCACCATTTGCTATGAACACGCGGTGGACA  
CGGTCACTCTACACATGTGGCCACATGTGCCTCTGCTACGCCTGTGGCCTGCGCCTCAAGAAGGCTCTGCA  
CGCCTGCTGCCCCATCTGCCGCCGCCCATCAAGGACATCATCAAGACCTACCGCAGCTCCTAG

**FIGURE 8C. Predicted amino acid sequence for the human homolog of CG11988 (557 amino acids); (SEQ ID NO:11)**

MGGQITRSTLHDSIGGPPFVTSRCHHKQKHCPAVLPSSGLPATPLLFPHTKGSQILMDLSHKAVKRQA  
SFCNAITFSNRPVLIYEQVRLKITKKQCCWSGALRLGFTSKDPSRIHPDSLPHYACPDVLSQSGFWAKAL  
PEEFANEGNIIAFWVDKGRVFRINDSAVMLFFSGVRTADPLWALVDVYGLTRGVQLLDSELVLPDCLR  
PRSFTALRRPSLRREADDARLSVSLCDLNVPGADGDEAAPAAGCPIPQNSLNSQHSRALPAQLDGLRFH  
ALRAGAHVRILDEQTVARVEHGRDERALVFTSRPVRVAETIFVKVTRSGGARPGALSFGVTTCDPGLRP  
ADLPFSPEALVDRKEFWAVCRVPGPLHSGDILGLVVDNADGELHLSHNGAAAGMQLCVDASQPLWMLFGLH  
GTITQIRILGSTILAERGIPSLPCSPASTPTSPSALGSRLSDPLLSTCSSGPLGSSAGGTAPNSPVSLE  
SPVTFPLGQWSDECTICYEHAVDTVIYTCGHMCLCYACGLRLKKALHACCPICRRPIKDIKTYRSS

**FIGURE 9.** Expression of neuralized-like in mammalian tissues – Real-time PCR analysis of neuralized-like expression in wildtype mouse tissues (DCT Pancreas = 23,34)



**FIGURE 10. HUMAN HOMOLOG OF CG8311****FIGURE 10A. BLASTP RESULTS FOR CG8311**

Homology to human gene ref NM\_014908.1; ref|NP\_055723.1| Protein

/protein=DTP06947034.1 /gene=DTG06947002.1 /locus=DTL06947020.1

/garid=G2R81HB986NMWM /chrom=9 /contig=NT\_023921.3

/start=112440 /end=114056 /strand=plus Similar to:

gi|7662482|ref|NP\_055723.1| KIAA1094 protein [Homo sapiens], Length = 1617

Score = 148 bits (371), Expect = 1e-35, Identities = 119/405 (29%), Positives = 189/405 (46%), Gaps = 19/405 (4%), Frame = +1

Query: 90 LTVAAGGMALETLCFFIYAFVKTGILVKCLVSLPGVATSLSFYLLVDTSLTFALIVGVFV 149  
+ VAA GMA+ + + + V L G+A + Y++ + +I  
Sbjct: 334 IVVAATGMAVALFSSVLALGITRPVPTNTCVIL--GLAGGVIIYIMKHSLSVGEVIEVLE 507

Query: 150 MTSAYQQIYIYTLRGFQRSFTYGEASVFVQGLVLFALSALHRLGGFFCGGSWPTEEFDTL 209  
+ + + + L R FT GEA + + G+ I R P + F +  
Sbjct: 508 VLLIFVYLLNMILLYLLPRCFTPEALLVLGGISFVLNQLIKRSLTLVESQGDVPDFFLLV 687

Query: 210 NMIMVNXXXXXXXXXXXXXXXXXPTLRKPCRFLWTVMLLLAVTCMP---VTRPXXXXXXX 265  
++ + T F+L T +L L V +P + R  
Sbjct: 688 VVGMVLMGIFFTLFFVMDSGTWASSIFFHLMTCVLSLGVV-LPWLHRLIRRNPLLWLL 864

Query: 266 XXXXRDOERLAILVFYMLLVLTCLTVAWQIGSSA-----KANTRVRKIFHLLIVMVY 318  
+ R+ +L ++ LL L CL V +Q + +A T RK FHL++V Y  
Sbjct: 865 QFLFQTDTRIYLLAYWSLLATLACLVLVLYQNAKRSSSESKKHQAPTIARKYFHLIVVATY 1044

Query: 319 IPGLIFECALLYLATGXXXXXXXXXXXXXXXXXKIPPFADRLAVAFSTFKDEKDAGELALTP 378  
IPG+IF+ LLY+A +I P L S F DE+D+G L LT  
Sbjct: 1045IPGIIFDRPLLYVAATVCLAVFIFLEYVRYFRIKPLGHTLRSFSLFLDERDSGPLILTH 1224

Query: 379 FCLLIGCSMPIWMTPCPCS-----GDNTLALLSGILAVGVGDTAASVVGSKLGRNKWGR 432  
LL+G S+PIW+ P PC+ G L +G+LAVGVGDT AS+ GS +G +W  
Sbjct: 1225IYLLLGMSLPWLIPRPCTQKGSGLGARALVPYAGVLAVGVGDTVASIFGSTMGEIRWPG 1404

Query: 433 SSRSLEGTTIAFVVSILMAVWLEI--SGLVAMSQAKWFATIFAALNSALVEAFTDQVDNL 490  
+ ++ EGT+ + + ++V L+ I SG+ W + + +L+EA+T Q+DNL  
Sbjct: 1405TKKTFEGTMTSIFAQIISVALILIFDSGVLDNYSYAWILGSISTV--SLLEAYTTQIDNL 1578

Query: 491 VLPL 494  
+LPL  
Sbjct: 1579LLPL 1590

**FIGURE 10B: Predicted coding sequence of the protein encoding the human CG8311 homolog (1617 bp), (SEQ ID NO:12)**

>DTT06947025

ATGACCCGAGAGTGGCCATCTCCGGCCCCGGGGCTGGGGCTCCGCTGAGTGGATCGGTGCTGGCAGAGG  
CGGAGTAGTGTGTGAGTGGTGTGAGCATCCACGCAACCGTATGGGACCGATACTCGTGGTGCGCCGT  
GGCCCTCGCAGTGCAGGCCCTTCTACGTCCAATACAAGTGGGACCGGCTGCTACAGCAGGGAAGCGCCGTC  
TTCCAGTTCGGAATGTCGCAACAGTGGCCTATTGCCCGCTCCATGGTTCATGCCCTTGCTTGGA  
TCATGAAGGAGCGGTGCCAGACTGCTGGGAACCGTTCCTTGAGCGTTTGGCATTGTGGTGGCAGCCAC  
TGGCATGGCAGTGGCCCTCTTCTCATCAGTGTGGCGCTCGGCATCACTCGCCAGTGCCAACCAACT  
TGTGTCATCTTGGGCTTGGCTGGAGGTGTTATCATTTATATCATGAAGCACTCGTTGAGCGTGGGGGAGG  
TGATCGAAGTCCTGGAAGTCTTCTGATCTTCGTTTATCTCAACATGATCCTGCTGTACCTGCTGCCCGG  
CTGCTTCACCCCTGGTGAGGCACTGCTGGTATTGGGTGGCATTAGCTTGTCTCTCAACAGCTCATCAAG

CGCTCTCTGACACTGGTGGAAAGTCAGGGGGACCCAGTGGACTTCTTCTGCTGGTGGTGGTAGTAGGGA  
TGGTACTCATGGGCATTTTCTTCAGCACTCTGTTTTGTCTTCATGGACTCAGGCACCTGGGGCCTCCTCCAT  
CTTCTTCCACCTCATGACCTGTGTGCTGAGCCTTGGTGTGGTCCCTACCCTGGCTGCACCGGCTCATCCGC  
AGGAATCCCCCTGCTCTGGCTTCTTCAGTTCTCTTCCAGACAGACACCCGCATCTACCTCCTAGCCTATT  
GGTCTCTGCTGGCCACCTTGGCCTGCCTGGTGGTGTGTACCAGAATGCCAAGCGGTCATCTTCCGAGTC  
CAAGAAGCACCAGGCCCCCACCATCGCCCCGAAAGTATTTCCACCTCATGTGGTAGCCACCTACATCCCA  
GGTATCATCTTTGACCGGCCACTGCTCTATGTAGCCGCCACTGTATGCCTGGCGGTCTTCATCTTCTCTGG  
AGTATGTGCGCTACTTCCGCATCAAGCCTTTGGGTCACTCTACGGAGCTTCTGTCCCTTTTCTGGA  
TGAACGAGACAGTGGACCACTCATTCTGACACACATCTACCTGCTCCTGGGCATGTCTCTTCCCCTCTGG  
CTGATCCCCAGACCCTGCACACAGAAGGGTAGCCTGGGAGGAGCCAGGGCCCTCGTCCCCTATGCCGGTG  
TCCTGGCTGTGGGTGTGGGTGATACTGTGGCCTCCATCTTCGGTAGCACCATGGGGGAGATCCGCTGGCC  
TGAACCAAAAAGACTTTTGGGGGACCATGACATCTATATTTGCGCAGATCATTCTGTAGCTCTGATC  
TTAATCTTTGACAGTGGAGTGGACCTAACTACAGTTATGCTTGGATTTGGGGTCCATCAGCACTGTGT  
CCCTCCTGGAAGCATACACTACACAGATAGACAATCTCTTCTGCCTCTCTACCTCCTGATATTGCTGAT  
GGCCTAG

**FIGURE 10C: Predicted amino acid sequence of the human CG8311 homolog protein (538 aa) (SEQ ID NO:13)**

>DTP06947034

MTRECPSPAPGPGAPLSGSLVLAEEAAVVFVAVLSIHATVWDRYSWCAVALAVQAFYVQYKWDRLQQGS  
FQFRMSANSGLLPASVMPLLLGLVMKERCQTAGNPFERFGIVVAATGMAVALFSSVLALGITRVPVPTNT  
CVILGLAGGVIIYIMKHSLSVGEVIEVLEVLIFVYLNMLILLYLLPRCFTPGEALLVLGGISFVLNQLIK  
RSLTLVESQGDVPDFFLLVVVGMVLMGIFSTLFFVMDSGTWASSIFHLMTCVLSLGVVLPWLHRLIR  
RNPLLWLLQFLFQTDTRIYLLAYWSLLATLACLVLVLYQNAKRSSSESXKHQAPTIAKYFHLIVVATYIP  
GTIFDRPLLYVAATVCLAVFIFLEYVRYFRICKPLGHTLRSFLSLFLDERDSGPLILTHIYLLGMSLPIW  
LIPRPTQKGS LGGARALVPYAGVLA VGVDTVASIFGSTMGEIRWPGTKKTFEGTMTSIFAQII SVALI  
LIFDSGVDLNYSAWILGSISTVSLLEAYTTQIDNLLLPLYLILLLMA

**FIGURE 10D: Transmembrane domains of the CG8311 homolog protein**

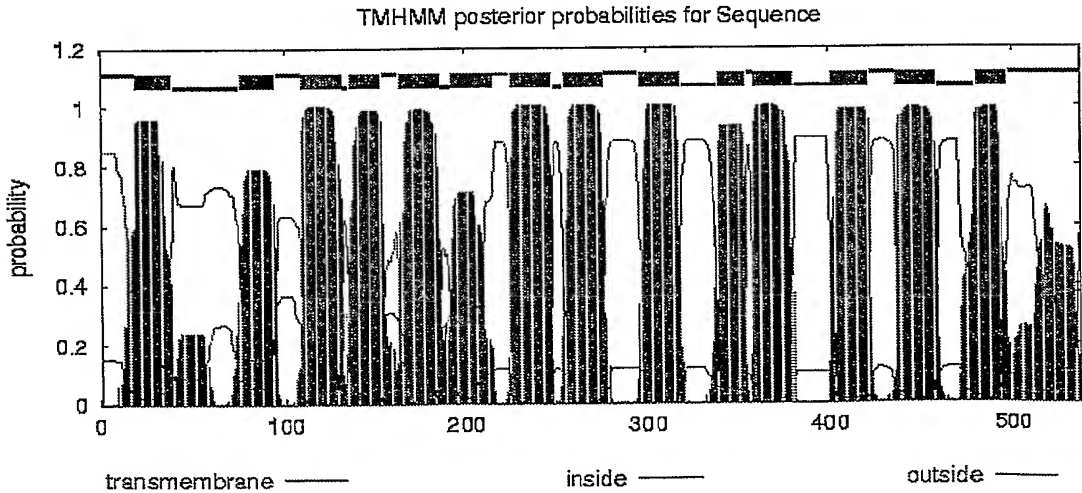
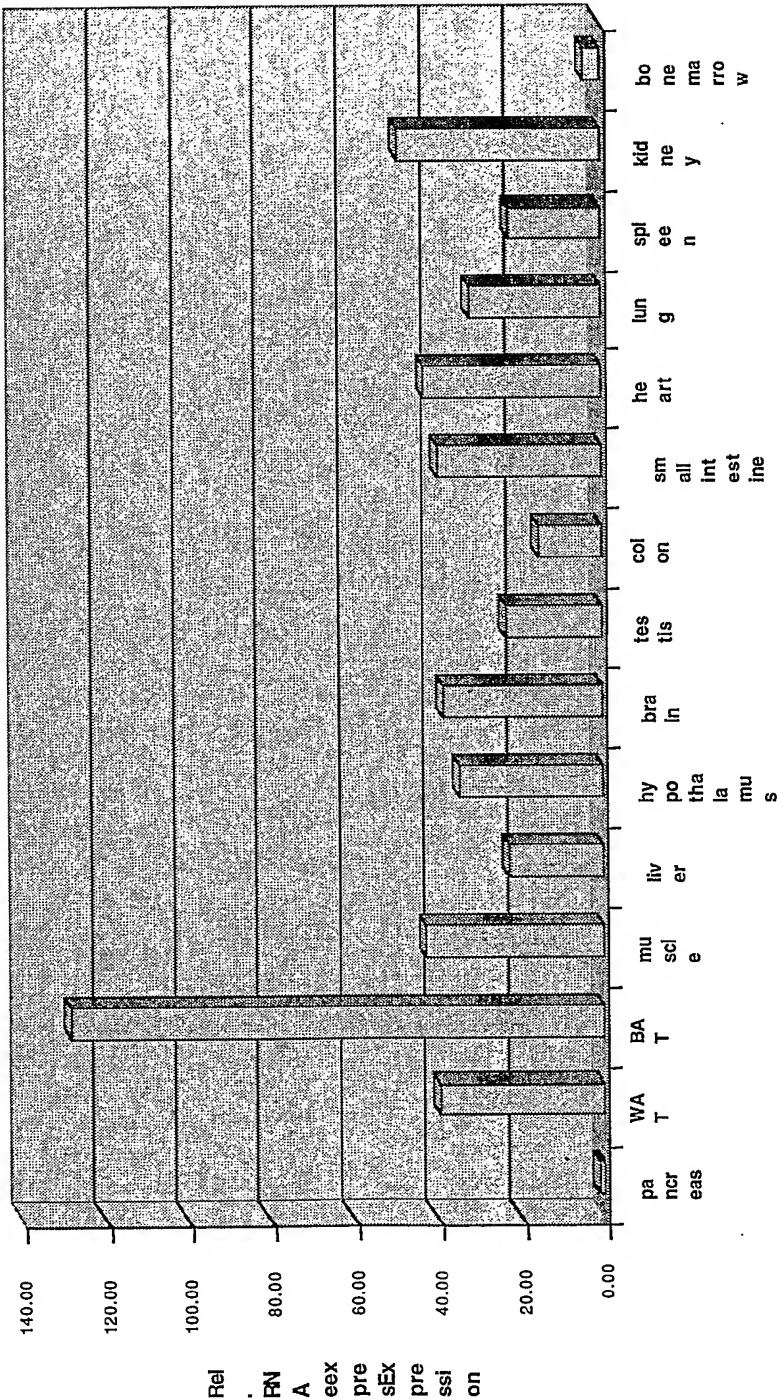


FIGURE 11: Expression of the CG8311 homolog in mammalian tissues - Real-time PCR analysis of CG8311 homolog expression in wildtype mouse tissues.



17/48

**FIGURE 12. A HUMAN HOMOLOG OF CG2048 (dco)****FIGURE 12A. BLASTP SEARCH RESULTS FOR CG2048**

Homology to human gene ref NM\_001893.1; ref NP\_001884.1 Protein

```

/protein=DTP10853018.1 /gene=DTG10853001.1 /locus=DTL10853006.1
  /garid=G2P4PMN52MHJGT /chrom=17 /contig=NT_025911.2
  /start=522 /end=23838 /strand=plus Similar to:
  gi|4503091|ref|NP_001884.1| casein kinase 1, delta [Homo
  sapiens], Length = 913

Score = 519 bits (1322), Expect(2) = e-150, Identities = 242/281 (86%), Positives
= 263/281 (93%), Frame = +2

```

```

Query: 15  IGSGSFGDIYLGTTINTGEEVAIKLECI RTKHPQLHIESKFYKTMQGGIGIPRIIWCGSE 74
          IGSGSFGDIYLG I GEEVAIKLEC++TKHPQLHIESK YK MQGG+GIP I WCG+E
Sbjct: 44  IGSGSFGDIYLGTDIAAGEEVAIKLECVKTKHPQLHIESKIYKMMQGGVGIGIPTIRWCGAE 223

Query: 75  GDYNVMVMELLGPSLEDLNFNCSRRFSLKTVLLADQMISRIDYIHSRDFIHRDIKPDNF 134
          GDYNVMVMELLGPSLEDLNFNCSR+FSLKTVLLADQMISRI+YIHS++FIHRD+KPDNF
Sbjct: 224 GDYNVMVMELLGPSLEDLNFNCSRKFSKTVLLADQMISRIEYIHSKNFIHRDVKPDNF 403

Query: 135 LMGLGKKGNLVYIIDFGLAKKFRDARSLKHIPYRENKNLTGTARYASINTHLGIEQSRRD 194
          LMGLGKKGNLVYIIDFGLAKK+RDAR+ +HIPYRENKNLTGTARYASINTHLGIEQSRRD
Sbjct: 404 LMGLGKKGNLVYIIDFGLAKKYRDARTHQHIPYRENKNLTGTARYASINTHLGIEQSRRD 583

Query: 195 DLESLGYVLMYFNLGALPWQGLKAANKRQKYERISEKKLSTSIIVLCKGFPSEFVNYLNF 254
          DLESLGYVLMYFNLG+LPWQGLKAA KRQKYERISEKK+ST I VLCKG+PSEF YLNF
Sbjct: 584 DLESLGYVLMYFNLGSLPWQGLKAATKRQKYERISEKKMSTPIEVLCKGYPSEFATYLN 763

Query: 255 CRQMHFDPQPDYCHLRKLFRLNLFHRLGFTYDYVFDWNLLKF 295
          CR + FD +PDY +LR+LFRNLFHR GF+YDYVFDWN+LKF
Sbjct: 764 CRSRLRFDDKPDYSYLRQLFRNLFHRQGFSDYDYVFDWNMLKF 886

```

```

Score = 31.3 bits (69), Expect(2) = e-150, Identities = 13/14 (92%), Positives =
14/14 (99%), Frame = +1

```

```

Query: 1  MELRVGNKYRLGRK 14
          MELRVGN+YRLGRK
Sbjct: 1  MELRVGNRYRLGRK 42

```

**FIGURE 12B. Predicted coding sequence for the human homolog with Accession Number NM\_001893.1 (Casein Kinase 1 delta) (913 bp), (SEQ ID NO:14)**

&gt;DTP10853009

```

atggagctgagagtcgggaacaggtaccggctgggcccgaagcatcggcagcggctccttcgggagacatc
tatctcgggtacggacattgctgcaggagaagaggttgccatcaagcttgaatgtgtcaaaaccaaacc
ctcagctccacattgagagcaaaatctacaagatgatgcaggagagtgggcatccccaccatcagatg
gtgcggggcagaggggactacaacgtcatggtgatggagctgctggggccaagcctggaggacctcttc
aacttctgctccaggaaattcagcctcaaaaccgtcctgctgcttgctgaccaaagatcagtcgcatcg
aatacattcattcaaagaacttcacccgggatgtgaagccagacaacttcctcatgggcctggggaa
gaagggcaacctggtgtacatcatcgacttcgggctggccaagaagtaccgggatgcacgcacccaccag
cacatccccctatcgtgagaacaagaacctcacggggacggcgcggtacgcctccatcaacacgcaccttg
gaattgaacaatcccgaagagatgacttggagctctctgggctacgtgctaattgtacttcaacctgggctc

```

tctccctggcaggggctgaaggctgccaccaagagacagaaatacgaaaggattagcgagaagaaaatg  
tccacccccatcgaagtgttggtgtaaaggctacccttccgaatttgccacatacctgaatttctgccgtt  
ccttgcgttttgacgacaagcctgactactcgctacctgcggcagcttttccggaatctgttccatcgcca  
gggcttctcctatgactacgtgttcgactggaacatgctcaaatttgtaagtcgcactgccagcacccggc  
tga

**FIGURE 12C. Predicted amino acid sequence for the human homolog with Accession Number NM\_001893.1 (Casein Kinase 1 delta) (304 aa), (SEQ ID NO:15)**

>DTP10853018

MELRVGNRYRLGRKHRQRLRRHLSXXTDIAAGEEVAIKLECVKTKHPQLHIESKIYKMMQGGVGIP TIR  
WCGAEGDYNVMVMELLGPSLEDLFNFC SRKFS LKTVLL LADQMISRIEYIHSKNFIHRDVKPDNFLMGLG  
KKG NLVYI IDFGLAKKYRDARTHQH I PYRENKNLTGTARYASINTHLGIEQSRDDLES LGYVLMYFNLG  
SLPWQGLKAATKRQKYERISEKKMSTPIEVLCKGYPSEFATYLNFCRSLRFDDKPDYSYLRQLFRNLFHR  
QGFSYDYVFDWNMLKFVSRTASTG

**FIGURE 13. A HUMAN HOMOLOG OF CG2048 (dco)****FIGURE 13A. BLASTP SEARCH RESULTS FOR CG2048**

Homology to human gene ref XM\_009983.4; ref XP\_009983.3 Protein

```
/protein=DTP12051036.1 /gene=DTG12051001.1 /locus=DTL12051024.1
  /garid=G2TW6QQ9W44RZ2 /chrom=22 /contig=NT_011520.5
  /start=17818584 /end=17839453 /strand=minus Similar to:
  gi|13655092|ref|XP_009983.2| casein kinase 1, epsilon
  [Homo sapiens], Length = 1535

Score = 462 bits (1177), Expect = e-130, Identities = 220/246 (89%), Positives =
236/246 (95%), Frame = +1
```

```
Query: 1 MELRVGNKYRLGRKIGSGSFGDIYLGTTINTGEEVAIKLECIIRTKHPQLHIESKFYKTMQ 60
      MELRVGNKYRLGRKIGSGSFGDIYLG I +GEEVAIKLEC++TKHPQLHIESKFYK MQ
Sbjct: 1 MELRVGNKYRLGRKIGSGSFGDIYLGANIASGEEVAIKLECVKTKHPQLHIESKFYKMMQ 180

Query: 61 GGIGIPRIIWCSEGDYNVMVMELLGPSLEDLNFNCSRFSLSKTVLLADQMISRIDYIH 120
      GG+GIP I WCG+EGDYNVMVMELLGPSLEDLNFNCSR+FSLKTVLLADQMISRI+YIH
Sbjct: 181 GGVGIPSIKWCAGEDYNVMVMELLGPSLEDLNFNCSRFSLSKTVLLADQMISRIEYIH 360

Query: 121 SRDFIHRDIKPDNFLMGLGKKGNLVYIIDFGLAKKFRDARSLKHIPYRENKNLTGTARYA 180
      S++FIHRD+KPDNFLMGLGKKGNLVYIIDFGLAKK+RDAR+ +HIPYRENKNLTGTARYA
Sbjct: 361 SKNFIHRDVKPDNFLMGLGKKGNLVYIIDFGLAKKYRDARTHQHIPPYRENKNLTGTARYA 540

Query: 181 SINTHLGIEQSRDDLESGLYVLMYFNLGALPWQGLKAANKRQKYERISEKKLSTSIIVL 240
      SINTHLGIEQSRDDLESGLYVLMYFNLG+LPWQGLKAA KRQKYERISEKK+ST I VL
Sbjct: 541 SINTHLGIEQSRDDLESGLYVLMYFNLGSLPWQGLKAATKRQKYERISEKKMSTPIEVL 720

Query: 241 CKGFPS 246
      CKG+PS
Sbjct: 721 CKGYPS 738
```

```
Score = 101 bits (249), Expect = 2e-21, Identities = 41/57 (71%),
Positives = 48/57 (83%), Frame = +3
```

```
Query: 245 PSEFVNYLNFRCRQMHFDQRPDYCHLRKLFRLNFHRLGFTYDYVFDWNLLKFGGPRNP 301
      P+EF YLNFRCR + FD +PDY +LR+LFRNLFHR GF+YDYVFDWN+LKFG RNP
Sbjct: 1017 PAEFSTYLNFCRSLRFDDKPDYSYLRQLFRNLFHRQGFSDYVFDWNMLKFGAARNP 1187
```

**FIGURE 13B: Predicted coding sequence for the human homolog with Accession Number XM\_009983.4 (Casein Kinase 1 epsilon) (1535 bp), (SEQ ID NO:16)**

>DTT12051027

```
ATGGAGCTACGTGTGGGGAACAAGTACCGCCTGGGACGGAAGATCGGGAGCGGGTCCTTCGGAGATATCT
ACCTGGGTGCCAACATCGCCTCTGGTGAGGAAGTCGCCATCAAGCTGGAGTGTGTGAAGACAAAGCACCC
CCAGCTGCACATCGAGAGCAAGTTCTACAAGATGATGCAGGGTGGCGTGGGGATCCCGTCCATCAAGTGG
TGCGGAGCTGAGGGCGACTACAACGTGATGGTCATGGAGCTGCTGGGGCCTAGCCTCGAGGACCTGTTCA
ACTTCTGTTCCTCCGCAAATTCAGCCTCAAGACGGTGCTGCTCTTGGCCGACCAGATGATCAGCCGCATCGA
GTATATCCACTCCAAGAACTTCATCCACCGGGACGTCAAGCCGACAACTTCCTCATGGGGCTGGGGAAG
AAGGGCAACCTGGTCTACATCATCGACTTCGGCCTGGCCAAGAAGTACCGGGACGCGCCGACCCACCAGC
ACATTCCTTACCGGGAAACAAGAACCTGACCGGCACGGCCCGCTACGCTTCATCAACACGCACCTGGG
CATTTGAGCAAAGCCGTCGAGATGACCTGGAGAGCCTGGGCTACGTGCTCATGTACTTCAACCTGGGCTCC
CTGCCCTGGCAGGGGCTCAAGCAGCCACCAAGCGCCAGAAGTATGAACGGATCAGCGAGAAGAAGATGT
```

CAACGCCCATCGAGGTCTCTGCAAAGGCTATCCCTCTGTGGGCATGGATGGCTCCTGGATCCGGCTACC  
TGGCAGAGGCTCCTGGCCACTGTTTCGCAGGTCTTCCCTTGCTCTGAAAGCAGCAGAGGCTCTGCTCACGG  
GCAGGCTAGGCTTTTCATCACAGTAGGATAGGGAAGGCCATGGCTCTGTTGCCTCCTTTTCTTGCTCTCA  
CAGATTGGAAGTATCTAGGGACAGTGGGTGGCTAGGACAGTGTGGCTGCAGGGGGTCTGGGAGCGTGGG  
CCTCACAGTGGCCTTCTCTATCCTCTGCAACACCTCCAGCCGAATTCTCAACATACCTCAACTTCTGCC  
GCTCCCTGCGGTTTGACGACAAGCCCGACTACTCTTACCTACGTCAGCTCTTCCGCAACCTCTTCCACCG  
GCAGGGCTTCTCCTATGACTACGTCTTTGACTGGAACATGCTGAAATTCGGTGCAGCCCGGAATCCCGAG  
GATGTGGACCGGGAGCGGCGAGAACACGAACGCGAGGAGAGGATGGGGCAGCTACGGGGGTCCGCGACCC  
GAGCCCTGCCCCCTGGCCCAACCCACGGGGGCCACTGCCAACCGCTCCGCAAGTCCGCGGAGCCCGTGGC  
TTCCACGCCAGCCTCCCGCATCCAGCCGGCTGGCAATACTTCTCCCAGAGCGATCTCGCGGGTCGACCGG  
GAGAGGAAGGTGAGTATGAGGCTGCACAGGGGTGCGCCCGCCAACGTCTCCTCCTCAGACCTCACTGGGC  
GGCAAGAGGTCTCCCGGATCCAGCCTCACAGACAAGTGTGCCATTTGACCATCTCGGGAAGTGA

**FIGURE 13C: Predicted amino acid sequence for the human homolog with Accession Number XM\_009983.4 (Casein Kinase 1 epsilon) (511 aa), (SEQ ID NO:17)**

>DTP12051036

MELRVGNKYRLGRKIGSGSFGDIYLGANIASGEEVAIKLECVKTKHPQLHIESKFYKMMQGGVGIPSIKW  
CGAEGDYNVMVMELLGPSLEDLFNFCSRKFSLKTVLLADQMISRIEYIHSKNFIHRDVKPDNFMGLGK  
KGNLVYIIDFGLAKKYRDARTHQHIPYRENKNLTGTARYASINTHLGIEQSRDDLES LGYVLMYFNLGS  
LPWQGLKAATKRQKYERISEKKMSTPIEVLCKGYPCGHGWLLDPATWQRLLATVRRSSLALKAAEALLT  
GRLGFHHSRIGKAMALLPPFLASHRLEVS RDSGWLGQCWLQGVWERGPHSGLLYPLQHPPAEFSTYLNFC  
RSLRFDDKPDYSYLRQLFRNLFRHQGFSYDYVFDWNMLKFGAARNPEDVDRE REREHEREERMGQLRGSAT  
RALPPGPPTGATANRLSAAEPVASTPASRIQPAGNTSPRAISRVDRE RKVSMRLHRGAPANVSSSDLTG  
RQEVSRIPASQTSVPFDHLGK

**FIGURE 13D: ClustaW alignment of Drosophila GadFly Accession Number CG2048 (referred to as 'dCK'), human casein kinase 1, delta (GenBank Accession Number NM\_001893.1; referred to as 'hCK delta'), human casein kinase 1, epsilon (GenBank Accession Number XM\_009983.4; referred to as 'hCK epsilon'), mouse casein kinase 1, delta (Accession Number AB028241.1; referred to as 'mCK delta'), mouse casein kinase 1, epsilon (Accession Number NM\_013767.2; referred to as 'mCK epsilon').**

dCK I	M E L R V G N K Y R L G R K I G S G S F G D I Y L G T T I N T G E E V A I K L E	40
hCK I delta	M E L R V G N R Y R L G R K I G S G S F G D I Y L G T D I A G E E E V A I K L E	40
hCK I epsilon	M E L R V G N K Y R L G R K I G S G S F G D I Y L G A N I A S G E E V A I K L E	40
mCK I delta	M E L R V G N K Y R L S R K I G S G S F G D I Y L G A N I A T G E E V A I K L E	40
mCK I epsilon	M E L R V G N K Y R L G R K I G S G S F G D I Y L G A D I A S G E E V A I K L E	40
dCK I	C I R T K H P D L H E S K E Y K T M Q G G I G T P R T I W C G S E G D Y N V M	80
hCK I delta	C V K T K H P D L H E S K I Y K M M Q G G V G I P T I R W C G A E G D Y N V M	80
hCK I epsilon	C V K T K H P D L H E S K F Y K M M Q G G V G I P S I K W C G A E G D Y N V M	80
mCK I delta	C V K T K H P D L H E S K F Y K M M Q G G V G I P S I K W C G A E G D Y N V M	80
mCK I epsilon	C V K T K H P D L H E S K F Y K M M Q G G V G I P S I K W C G A E G D Y N V M	80
dCK I	V M E L L G P S L E D L F N F C S R R F S L K T V L L L A D Q M I S R I E Y I H	120
hCK I delta	V M E L L G P S L E D L F N F C S R K F S L K T V L L L A D Q M I S R I E Y I H	120
hCK I epsilon	V M E L L G P S L E D L F N F C S R K F S L K T V L L L A D Q M I S R I E Y I H	120
mCK I delta	V M E L L G P S L E D L F N F C S R K F S L K T V L L L A D Q M I S R I E Y I H	120
mCK I epsilon	V M E L L G P S L E D L F N F C S R K F S L K T V L L L A D Q M I S R I E Y I H	120
dCK I	S R D F I H R D V I K P D N F L M G L G K K G N L V Y I I D F G L A K K Y R D A R	160
hCK I delta	S K N F I H R D V I K P D N F L M G L G K K G N L V Y I I D F G L A K K Y R D A R	160
hCK I epsilon	S K N F I H R D V I K P D N F L M G L G K K G N L V Y I I D F G L A K K Y R D A R	160
mCK I delta	S K N F I H R D V I K P D N F L M G L G K K G N L V Y I I D F G P G K K Y R D A R	160
mCK I epsilon	S K N F I H R D V I K P D N F L M G L G K K G N L V Y I I D F G L A K K Y R D A R	160
dCK I	S L K H I P Y R E N K N L T G T A R Y A S I N T H L G I E Q S R R D D L E S L G	200
hCK I delta	T H C H I P Y R E N K N L T G T A R Y A S I N T H L G I E Q S R R D D L E S L G	200
hCK I epsilon	T H C H I P Y R E N K N L T G T A R Y A S I N T H L G I E Q S R R D D L E S L G	200
mCK I delta	T H C H I P Y R E N K N L T G T A R Y A S I N T H L G I E Q S R R D D L E S L G	200
mCK I epsilon	T H C H I P Y R E N K N L T G T A R Y A S I N T H L G I E Q S R R D D L E S L G	200
dCK I	Y V L M Y F N L G S L P W Q G L K A A T K R Q K Y E R I S E K K L S T S I V V L	240
hCK I delta	Y V L M Y F N L G S L P W Q G L K A A T K R Q K Y E R I S E K K M S T P I E V L	240
hCK I epsilon	Y V L M Y F N L G S L P W Q G L K A A T K R Q K Y E R I S E K K M S T P I E V L	240
mCK I delta	Y V L M Y F N L G S L P W Q G L K A A T K R Q K Y E R I S E K K M S T P I E V L	240
mCK I epsilon	Y V L M Y F N L G S L P W Q G L K A A T K R Q K Y E R I S E K K M S T P I E V L	240
dCK I	C K G Y P S E F V N Y L N F C R Q M H F D Q R P D Y C H L R K L F R N L F H R Q	280
hCK I delta	C K G Y P S E F A T Y L N F C R S L R F D D K P D Y S Y L R Q L F R N L F H R Q	280
hCK I epsilon	C K G Y P S E F S T Y L N F C R S L R F D D K P D Y S Y L R Q L F R N L F H R Q	280
mCK I delta	C K G Y P S E F S T Y L N F C R S L R F D D K P D Y S Y L R H V F R N L F H R Q	280
mCK I epsilon	C K G Y P S E F S T Y L N F C R S L R F D D K P D Y S Y L R Q L F R N L F H R Q	280
dCK I	G F T Y D Y V F D W N L L K F G G P R N R Q A I Q Q A Q D I S A D G Q A G H D A V	320
hCK I delta	G F S Y D Y V F D W N L L K F G A S R A A D D A E R E R R I D - - R E E R L R H S	318
hCK I epsilon	G F S Y D Y V F D W N L L K F G A A R N P E D V D R E R R E H E R E E R M G Q L	320
mCK I delta	G F S Y D Y V F D W N L L K F G A A R N P E D V D R E R R E H E R E E R M G Q L	320
mCK I epsilon	G F S Y D Y V F D W N L L K F G A A R N P E D V D R E R R G H E R E E R M G Q L	320
dCK I	A A A A A V A A A A A S S H Q Q Q Q H K V N A A L G G G G G S A A Q Q Q L Q G	360
hCK I delta	R N P A T R G L P S - - - - T D S G R L R G T Q E V A P P T P L T P T S H T	352
hCK I epsilon	R G S A T R A L P P G P P T - - G A T A N R L R S A A E P V A S T P A S R I D P A	359
mCK I delta	R G S A T R A L P P G P P T - - G A T A N R L R S A A E P V A S T P A S R I Q Q T	359
mCK I epsilon	R G S A T R A L P P G P P T - - G A T A N R L R S A A E P V A S T P A S R I Q Q T	359
dCK I	G Q T L A M L G G N G G G N G S Q L I G G N G L N M D D S M A A T N S S R P P Y	400
hCK I delta	A N T S P R - P V S G M E R E R K V - - - - S M R L H R G A P V N I S S S - -	384
hCK I epsilon	G N T S P R - A I S R A V D R E R K V - - - - S M R L H R G A P A N V S S S - -	391
mCK I delta	G N T S P R - A I S R A D R E R K V - - - - S M R L H R G A P A N V S S S - -	391
mCK I epsilon	G N T S P R - A I S R A D R E R K V - - - - S M G L H R G A P A N V S S S - -	391
dCK I	D T P E R R P S I R M R Q G G G G G G G G V G V G G M P S G G G G G G V G N A K	440
hCK I delta	D L T G R Q D T S R M - - S T S Q I P G R V A S S G L Q S V V - - - - H R	415
hCK I epsilon	D L T G R Q E V S R I - - P A S Q T S V P F D H L G - - - - - - - - K	416
mCK I delta	D L T G R Q E V S R I - - A A S Q T S V P F D H L G - - - - - - - - K	416
mCK I epsilon	D L T G R Q E V S R I - - A A S Q T S V P F D H L G - - - - - - - - K	416

Decoration 'Decoration #1': Shade (with solid bright yellow) residues that match the Consensus exactly.

**FIGURE 14: A HUMAN HOMOLOG OF CG5320 (Gdh)****FIGURE 14A: BLASTP SEARCH RESULTS FOR CG5320**

Homology to human gene ref NM\_005271; ref NP\_005262 Protein

/protein=DTP12372021.1 /gene=DTG12372001.1 /locus=DTL12372010.1  
 /garid=G2TFM0S\_5SF0QKH /chrom=X /contig=NT\_011746.3  
 /start=793092 /end=794768 /strand=minus Similar to:  
 gi|4885281|ref|NP\_005262.1| glutamate dehydrogenase 1 [Homo sapiens]  
 Length = 1677

Score = 749 bits (1913), Expect = 0.0, Identities = 367/553 (66%), Positives = 435/553 (78%), Frame = +1

Query: 12 ARRQELATLAKALPTAVMQSSRGYATEHQIPDRLKDVPTAKDPRFFDMVEYFFHRCQI 71  
 A L + P A Q A + + D DP FF MVE FF RG I  
 Sbjct: 67 ANHSAALLGRGRGQPAAASQPGLALAARRHYSELVAD--REDDEPNFFKMVEGFFDRGASI 240

Query: 72 AEESLVDDMKGLTRDEKKQKVKGILMLMQPCDHIIIEIAFPLRRDAGNYEMITGYRAQHS 131  
 E+ LV D++ + + ++K+ +V+GIL +++PC+H++ ++FP+RRD G++E+I GYRAQHS  
 Sbjct: 241 VEDKLVKDLRTQESEEQKRNRVRGILRIIKPCNHVLSLSFPIRRDDGSWEVIEGYRAQHS 420

Query: 132 THKTPTKGGKCFIRFSLDVSDEVKALSALMTFKCACVDVPFGGAKAGLKINPKEYSEHEL 191  
 H+TP KGG IR+S DVS DEVKAL++LMT+KCA VDVPFGGAKAG+KINPK Y+E+EL  
 Sbjct: 421 QHRTPCCKG--IRYSTDVSVDEVKALASLMTYKCAVVDVPFGGAKAGVKINPKNYTENEL 594

Query: 192 EKITRRFTLELAKKGFIFGPGVDVPAPDMGTGEREMSWIADTYAKTIGHLDINAHACVTGK 251  
 EKITRRFT+ELAKKGFIFGPGVDVPAPDM TGEREMSWIADTYA TIGH DINAHACVTGK  
 Sbjct: 595 EKITRRFTMELAKKGFIFGPGVDVPAPDMNTGEREMSWIADTYASTIGHYDINAHACVTGK 774

Query: 252 PINQGGIIGHRVSATGRGVFHGLENFINEANYMSQIGTTPGWGGKTFIVQGFNVGLHTTR 311  
 PI+QGGIIGHR+SATGRGVFHG+ENFINEA+YMS +G TPG+ KTF+VQGFNVGLH+ R  
 Sbjct: 775 PISQGGIIGHRISATGRGVFHGIENFINEASYMSILGMTPGFRDKTFVQGFNVGLHSMR 954

Query: 312 YLTRAGATCIGVIEHDGTLYNPEGIDPKLLEDYKNEHGTIVGYQNAKPYEGENLMFEKCD 371  
 YL R GA CI V E DG+++NP+GIDPK LED+K +HG+I+G+ AKPYEG L + CD  
 Sbjct: 955 YLHRFGAKCIAVGESDGSIWNPDGIDPKELEDFKLQHGSI LGFPKAKPYEGSILEVD-CD 1131

Query: 372 IFIPAAVEKVITSENANRIQAKIIAEANGPTTPAADKILIDRNILVIPDLYINAGGVTV 431  
 I IPAA EK +T NA R++AKIIAE ANGPTTP ADKI ++RNILVIPDLY+NAGGVTV  
 Sbjct: 1132ILIPAAATEKQLTKSNAPRVKAKIIAEGANGPTTPEADKIFLERNILVIPDLYLNAGGVTV 1311

Query: 432 SFFEWLKNLNHVSYGRLTFKYERESNYHLLASVQQSIERIINDESQESLERRFRGVGGR 491  
 S+FEWLKNLNHVSYGRLTFKYER+SNYHLL SVQESLER+FG+ GG  
 Sbjct: 1312SYFEWLKNLNHVSYGRLTFKYERDSNYHLL-----SVQESLERKFGKHGGT 1452

Query: 492 IPVTPSESFQKRISGASEKDIVHSGLDYTMERSARAIMKTAMKYNLGLDLRTAAYVNSIE 551  
 IP+ P+ FQ ISGASEKDIVHS L YTMERSAR IM TAMKYNLGLDLRTAAYVN+IE  
 Sbjct: 1453IPIVPTAEFQDSISGASEKDIVHSALAYTMERSARQIMHTAMKYNLGLDLRTAAYVNAIE 1632

Query: 552 KIFTTYRDAGLAF 564  
 K+F Y +AG+ F  
 Sbjct: 1633KVFKVYSEAGVTF 1671

**FIGURE 14B: Predicted coding sequence for the human homolog with Accession Number NM\_005271.1 (Glutamate dehydrogenase I) (1677 bp), (SEQ ID NO:18)**

>DTT12372012

```
ATGTACCGCTACCTGGCCAAAGCGCTGCTGCCGTCCCGGGCCGGGCCCCGCTGCCCTGGGCTCCGCGGCCA
ACCACTCGGCCCGCTTGCTGGGCCGGGGCCGCGGACAGCCCGCCGCCCTCGCAGCCGGGGCTCGCATT
GGCCGCCCCGGCGCCACTACAGCGAGTTGGTGGCCGACC GCGAGGACGACCCCAACTTCTTCAAGATGGTG
GAGGGCTTCTTCGATCGCGGCCAGCATCGTGGAGGACAAGTTGGTGAAGGACCTGAGGACCCAGGAAA
GCGAGGAGCAGAAGCGGAACCGGGTGC GCGGCATCCTGCGGATCATCAAGCCCTGCAACCATGTGCTGAG
TCTCTCCTTCCCCATCCGGCGCGACGACGGCTCCTGGGAGGTCATCGAAGGCTACCGGGCCCAGCACAGC
CAGCACCGCACGCCCTGCAAGGGAGGTATCCGTTACAGCACTGATGTGAGTGTAGATGAAGTAAAAGCTT
TGGCTTCTCTGATGACATACAAGTGTGAGTGGTTGATGTGCCGTTTGGGGGTGCTAAAGCTGGTGTAA
GATCAATCCCAAGAAGTATACCGAAAATGAATTGGAAAAGATCACAAGGAGGTTACACCATGGAGCTAGCA
AAGAAGGGCTTTATTGGTCTGGCGTTGATGTGCCCTGCTCCAGACATGAACACAGGTGAGCGGGAGATGT
CCTGGATGTGCTGATACCTATGCCAGCACCATAGGGCACTATGATATTAATGCACACGCCTGTGTTACTGG
TAAACCCATCAGCCAAGGGGGAATCCATGGACGCATCTCTGCTACTGGCCGTGGTGTCTTCCATGGGATT
GAAAACCTTCATCAATGAAGCTTCTTACATGAGCATTTTAGGAATGACACCAGGGTTTAGAGATAAAACAT
TTGTTGTTTCAGGGATTTGGTAATGTGGGCCCTACACTCTATGAGATATTTACATCGTTTTTGGTGTAAATG
TATTGCTGTTGGTGAGTCTGATGGGAGTATATGGAATCCAGATGGTATTTGACCCAAAGGAACTGGAAGAC
TTCAAATTGCAACATGGGTCCATTCTGGGCTTCCCCAAGGCAAAGCCCTATGAAGGAAGCATCTTGGAGG
TCGACTGTGACATACTGATCCCAGCTGCCACTGAGAAGCAGTTGACCAAATCCAACGCACCCAGAGTCAA
AGCCAAGATCATTGCTGAAGGTGCCAATGGGCCAACAACTCCAGAAGCTGATAAGATCTTCTGGAGAGA
AACATTTTGGTTATTCCAGATCTCTACTTGAATGCTGGAGGAGTGACAGTATCTTACTTTGAGTGGCTGA
AGAATCTAAATCATGTGAGCTATGGCCGTTTGACCTTCAAATATGAAAGGGATTCTAACTACCACTTGCT
CCTGTCTGTTCAAGAGAGTTTAGAAAGAAAATTTGGAAAGCATGGTGAAGTATTTCCATTGTACCCACG
GCAGAGTTCCAAGACAGTATATCGGGTGCATCTGAGAAAGACATTTGTGCACTCTGCCTTGGCATAACAAA
TGGAGCGTTCTGCCAGGCAAATATGCACACAGCCATGAAGTATAACCTGGGATTGGACCTGAGAACAGC
TGCCTATGTCAATGCCATTGAAAAAGTCTTCAAAGTGACAGTGAAGCTGGTGTGACCTTCACATAG
```

**FIGURE 14C: Predicted amino acid sequence for the human homolog with Accession Number NM\_005271.1 (Glutamate dehydrogenase I) (558 aa), (SEQ ID NO:19)**

>DTP12372021

```
MYRYLAKALLPSRAGPAALGSAANHSAALLGRGRGQPAAASQPLALAAARRHYSELVADREDDPNFFKMV
EGFFDRGASIVEDKLVKDLRTQESEEQKRNVRVGLIRIIKPCNHVLSLSFPIRRDDGSWEVIEGYRAQHS
QHRTPCCKGIRYSTDVSVDEVKALASLMTYKCAVVDVPPGGAKAGVKINPKNYTENELEKITRRFTMELA
KKGFIGPGVDVPAPDMNTGEREMSWIADTYASTIGHYDINAHACVTGKPI SQGGIHGRISATGRGVFHGI
ENFINEASYMSILGMTPGFRDKTFVVQGFVGNVGLHSMRYLHREFGAKCIAVGESDGSIWNPDGIDPKELED
FKLQHGSILGFPKAKPYEGSILEVDCDILIPAATEKQLTKSNAPRVKAKIIAEGANGPTTPEADKIFLER
NILVIPDLYLNAGGVTVSYFEWLKLNHVSYGRITFKYERDSNYHLLLSVQESLERKFGKHGGTIPIVPT
AEFQDSISGASEKDIVHSALAYTMERSARQIMHTAMKYNLGLDLRTAAYVNAIEKVFKVYSEAGVTFT
```

**FIGURE15: A HUMAN HOMOLOG OF CG5320 (Gdh)****FIGURE 15A: BLASTP SEARCH RESULTS FOR CG5320**

Homology to human gene ref NT\_011746.5

/garid=G4JMN1P\_524L23J /chrom=X /contig=NT\_011746.5 /start=1803270 /end=1807726  
/strand=minus GLUD2: Glutamate dehydrogenase-2, Length = 3920

Score = 749 bits (1913), Expect = 0.0, Identities = 367/553 (66%), Positives = 435/553 (78%), Frame = +2

Query: 12 ARRQQLATLAKALPTAVMQSSRGYATEHQIPDRLKDVPTAKDPRFFDMVEYFFHRCQI 71  
A L + P A Q A + + D DP FF MVE FF RG I  
Sbjct: 278 ANHSAALLGRGRGQPAAASQPGLALAARRHYSELVAD--REDDPNFFKMVEGFFDRGASI 451

Query: 72 AEESLVDDMKGLTRDEKKQKVKGILMLMQPCDHIIIEIAFFLRRDAGNYEMITGYRAQHS 131  
E+ LV D++ + + ++K+ +V+GIL +++PC+H++ ++FP+RRD G++E+I GYRAQHS  
Sbjct: 452 VEDKLVKDLRTQESEEQKRNRVRGILRIIKPCNHVLSLSFPPIRRDDGSWEVIEGYRAQHS 631

Query: 132 THKTPTKGGKCIRFSLDVSDEVKALSALMTFKCACVDVPFGGAKAGLKINPKEYSEHEL 191  
H+TP KGG IR+S DVS DEVKAL++LMT+KCA VDVFPFGGAKAG+KINPK Y+E+EL  
Sbjct: 632 QHRTFCKGG--IRYSTDVSDEVKALASLMTYKCAVVDVPFGGAKAGVKINPKNYTENEL 805

Query: 192 EKITRRFTLELAKKGFIPGVDPVPAPDMGTGEREMSWIADTYAKTIGHLDINAHACVTGK 251  
EKITRRFT+ELAKKGFIPGVDPVPAPDM TGEREMSWIADTYA TIGH DINAHACVTGK  
Sbjct: 806 EKITRRFTMELAKKGFIPGVDPVPAPDMNTGEREMSWIADTYASTIGHYDINAHACVTGK 985

Query: 252 PINQGGIHDRVSATGRGVFPHGLENFINEANYMSQIGTTPGWGGKTFIVQGFQGNVGLHTTR 311  
PI+QGGIHDR+SATGRGVFPHG+ENFINEA+YMS +G TPG+ KTF+VQGFQGNVGLH+ R  
Sbjct: 986 PISQGGIHDRISATGRGVFPHGIENFINEASYMSILGMTPGFRDKTFVVQGFQGNVGLHSMR 1165

Query: 312 YLTRAGATCIGVIEHDGTLNPEGIDPKLLEDYKNEHGTIVGYQNAKPYEGENLMFEKCD 371  
YL R GA CI V E DG+++NP+GIDPK LED+K +HG+I+G+ AKPYEG L + CD  
Sbjct: 1166YLHRFGAKCIAVGESDGSIWNPDGIDPKELEDFKLQHGSILGFPKAKPYEGSILEVD-CD 1342

Query: 372 IFIPAAVEKVITSENANRIQAKIIAEEANGPTTPAADKILIDRNILVIPDLYINAGGVTV 431  
I IPAA EK +T NA R++AKIIAE ANGPTTP ADKI ++RNILVIPDLY+NAGGVTV  
Sbjct: 1343ILIPAATEKQLTKSNAPRVKAKIIAEGANGPTTPEADKIFLERNILVIPDLYLNAGGVTV 1522

Query: 432 SFFEWLKNLNHVSYGRLTFKYERESNYHLLASVQQSIERIINDESVQESLERRFGRVGGR 491  
S+FEWLKNLNHVSYGRLTFKYER+SNYHLL SVQESLER+FG+ GG  
Sbjct: 1523SYFEWLKNLNHVSYGRLTFKYERDSNYHLL-----SVQESLERKFGKHGGT 1663

Query: 492 IPVTPSESFQKRISGASEKDIVHSGLDYTMERSARAIMKTAMKYNLGLDLRTAAYVNSIE 551  
IP+ P+ FQ ISGASEKDIVHS L YTMERSAR IM TAMKYNLGLDLRTAAYVN+IE  
Sbjct: 1664IPIVPTAEFQDSISGASEKDIVHSALAYTMERSARQIMHTAMKYNLGLDLRTAAYVNAIE 1843

Query: 552 KIFTTYRDAGLAF 564  
K+F Y +AG+ F  
Sbjct: 1844KVFKVYSEAGVTF 1882

**FIGURE 15B: Predicted coding sequence for the human homolog with Accession Number NT\_011746.5 (Glutamate dehydrogenase II) (1665 bp), (SEQ ID NO:20)**

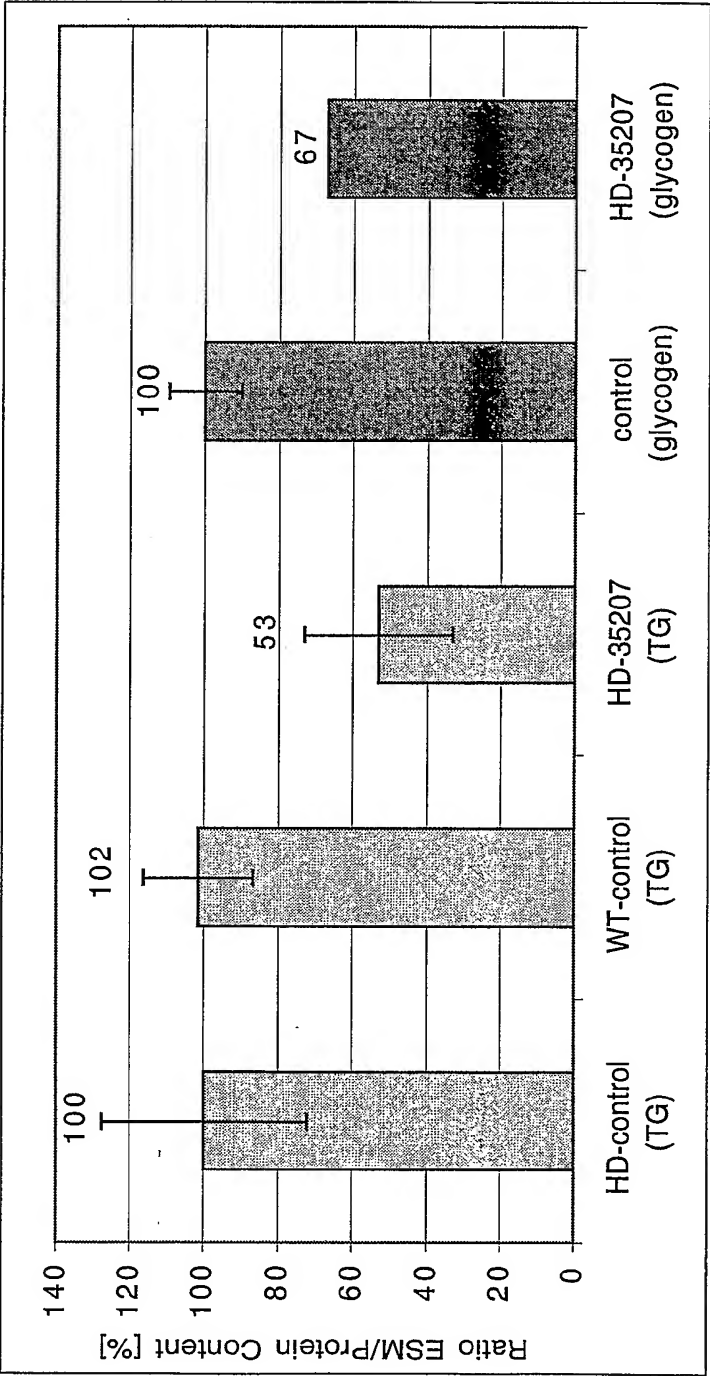
>DTT12372012.2 NT\_011746.5:

```
ATGTACCGCTACCTGGCCAAAGCGCTGCTGCCGTCCCGGGCCGGGCCCCGCTGCCCTGGGCTCCGCGGCCA
ACCACTCGGCCCGCTTGTGGCCCGGGCCGCGGACAGCCCCGCCGCCCTCGCAGCCGGGGCTCGCATT
GGCCGCCCGGCCCACTACAGCGAGTTGGTGGCCGACCGCGAGGACGACCCCAACTTCTTCAAGATGGTG
GAGGGCTTCTTCGATCGCGGCCAGCATCGTGGAGGACAAGTTGGTGAAGGACCTGAGGACCCAGGAAA
GCGAGGAGCAGAAGCGGAACCGGTGCGCGGCATCCTGCGGATCATCAAGCCCTGCAACCATGTGCTGAG
TCTCTCCTTCCCCATCCGGCGCGACGACGGCTCCTGGGAGGTATCGAAGGCTACCGGGCCAGCACAGC
CAGCACCGCACGCCCTGCAAGGGAGGTATCCGTTACAGCACTGATGTGAGTGTAGATGAAGTAAAAGCTT
TGGCTTCTCTGATGACATACAAGTGTGCAGTGGTTGATGTGCCGTTTGGGGGTGCTAAAGCTGGTGTAA
GATCAATCCCAAGAACTATACCGAAAATGAATTGGAAAAGATCACAAGGAGGTTTACCATGGAGCTAGCA
AAGAAGGGCTTTATTGGTCTTGGCGTTGATGTGCCTGCTCCAGACATGAACACAGGTGAGCGGGAGATGT
CCTGGATTGCTGATACCTATGCCAGCACCATAGGGCACTATGATATTAATGCACACGCCCTGTGTACTGG
TAAACCATCAGCCAAGGGGGAATCCATGGACGCATCTCTGCTACTGGCCGTGGTGTCTTCCATGGGATT
GAAAACCTTCATCAATGAAGCTTCTTACATGAGCATTTTAGGAATGACACCAGGGTTTAGAGATAAAACAT
TTGTTGTTTCAAGGATTGTTGTAATGTGGGCCTACACTCTATGAGATATTTACATCGTTTTGGTGTAAATG
TATTGCTGTTGGTGAAGTCTGATGGGAGTATATGGAATCCAGATGGTATTGACCCAAAGGAAC'TGGAAGAC
TTCAAATTGCAACATGGGTCCATTCTGGGCTTCCCCAAGGCAAAGCCCTATGAAGGAAGCATCTTGGAGG
TCGACTGTGACATACTGATCCAGCTGCCACTGAGAAGCAGTTGACCAAATCCAACGCACCCAGAGTCAA
AGCCAAGATCATTGCTGAAGGTGCCAATGGGCCAACAACTCCAGAAGCTGATAAGATCTTCCCTGGAGAGA
AACATTTTGGTTATTCCAGATCTCTACTTGAATGCTGGAGGAGTGACAGTATCTTACTTTGAGTGGCTGA
AGAATCTAAATCATGTGAGCTATGGCCGTTTACCTTCAAATATGAAAGGGATTCTAACTACCCTTGGCT
CCTGTCTGTTCAAGAGAGTTTAGAAAGAAAATTTGAAAGCATGGTGGAACTATTCCTATTGTACCCACG
GCAGAGTTCCAAGACAGTATATCGGGTGCATCTGAGAAAGACATTGTGCACTCTGCCTTGGCATACACAA
TGGAGCGTTCTGCCAGGCAAATTATGCACACAGCCATGAAGTATAACCTGGGATTGGACCTGAGAACAGC
TGCCTATGTCAATGCCATTGAAAAAGTCTTCAAAGTGTACAGTGAAGCTGTGTGA
```

**FIGURE 15C: Predicted amino acid sequence for the human homolog with Accession Number NT\_011746.5 (554 aa), (SEQ ID NO:21)**

```
MYRYLAKALLPSRAGPAALGSAANHSAALLGRGRGQPAASQPGLALAARRHYSELVADREDDPNFFKMV
EGFFDRGASIVEDKLVDLRTQEESEEQKRNVRVIGILRIKPCNHVLSLSFPIRRDDGSWEVIEGYRAQHS
QHRTPCCKGGIRYSTDVSVDEVKALASLMTYKCAVVDVPFGGAKAGVKINPKNYTENELEKITRRFTMELA
KKGFIGPVDVPAPDMNTGEREMSWIADTYASTIGHYDINAHACVTGKPI SQGGIHGRISATGRGVFHGI
ENFINEASYMSILGMTPGFRDKTFVVQGFNGVGLHSMRYLHRFGAKCIAVGESDGSIWNPDGIDPKELED
FKLQHGSI LGFPKAKPYEGSILEVDCDILIPAATEKQLTKSNAPRVKAKIIAEGANGPTTPEADKIFLER
NILVIPDLVYNAGGVTVSYFEWLKLNHVSYGRLTFKYERDSNYHLLLSVQESLERKFGKHGGTIPIVPT
AEFQDSISGASEKDIVHSALAYTMERSARQIMHTAMKYNLGLDLRTAAYVNAIEKVFVKVYSEAV
```

FIGURE 16. Energy storage metabolite content of a Drosophila Gdh (Gadfly Acc. No. CG5320) mutant



**FIGURE 17. HUMAN HOMOLOG OF CG3943 (kraken)****FIGURE 17A. tBLASTN SEARCH RESULT FOR CG3943**

Homology to human gene with GenBank Accession Number AL450314

dtgic|HSC140179.1 Identical to: Novel human gene mapping to chromosome 22.

Score = 149 bits (372), Expect = 3e-35, Identities = 99/289 (34%),  
Positives = 158/289 (54%), Gaps = 11/289 (3%), Frame = -1

```

Query: 41      EFSIAVPWGTVEAKWWSKERQPIIALHGWQDNCGSFDRLCPLLPADTSILALDLPGHGK 100
              E +AVPWG + AK WGS + P++ LHGW DN SFDRL PLLP D +A+D GHG
Sbjct: 1255    ELKLAVPWGHIAAKAWGSLQGPPVLCILHGWLNDASSFDRLLIPLLPQDFYYVAMDFGGHGL 1076

Query: 101     SSHYPMGMQYFIFWDGICLIIRIVRKYNWKNVTLLGHSLLGGALTFMYAASFPEVEKLIN 160
              SSHY G+ Y++ + IRR+V W ++LGHS GG + M+ +FP V+KLI
Sbjct: 1075    SSHYSPGVPPYYL-QTFVSEIRRVAALKWNRFSILGHSFSGGVVGMFFCTFPEMVDKLIL 899

Query: 161     IDIAGPTVRG--TQRMAGTGRALDKFLDYETLPESKQPCYSYDEMIKLVLDAYDGSVDE 218
              +D + + + RA++ L E E +S +++ +L + + + E
Sbjct: 898     LDTPLFLLESDEMENLLTYKRRRAIEHVLQVEASQEP SH-VFSLKQLLQRLKLS-NSHLSE 725

Query: 219     PSVRVLMNRGMHRNPSKNGYLFARDLRLKVSLLGM-FTAEQTLAYA-RQIRCRVLNIRGI 276
              +L+ RG G + RD RL + + F + + A++ R+++ VL I+ +
Sbjct: 724     ECGELLQQRGT--TKVATGLVLNRDQLAWAENSIDFISRELCAHSIRKLQAHVLLIKAV 551

Query: 277     PGMKFETPQVYAD-----VIATLREN-AAKVYVVEVPGTHHLHLVTPDRVAPHIIRFLK 329
              G F++ Q Y++ +I T++ + +VEVPG H +H+ P VA I FL+
Sbjct: 550     HG Y-FDSRQNYSEKESLSFMIDTMKSTLKEQFQFVEVPGNHCVMHSEPHVASIISFLQ 374

```

**FIGURE 17B. Predicted coding sequence for the human homolog of CG3943 (945 base pairs), (SEQ ID NO:22)**

```

ATGAGTGAGAACGCCGACACAGGCTCTGATCTCAGAGCTGAAGCTGGCTGTGCCCTGGGGCCACATCGCAGCCAAAGCCTGGGG
CTCCCTGCAGGGCCCTCCAGTTCTCTGCCTGCACGGCTGGCTGGACAATGCCAGCTCCTTCGACAGACTCATCCCTCTTCTCC
CGCAAGACTTTTATTACGTTGCCATGGATTTCGGAGGTCATGGGCTCTCGTCCCATTACAGCCCAGGTGTCCCATATTACCTC
CAGACTTTTGTGAGTGAGATCCGAAGAGTTGTGGCAGCCTTGAAATGGAATCGATTCTCCATTCTGGGCCACAGCTTCGTTGG
CGTCGTGGGCGGAATGTTTTCTGTACCTTCCCGAGATGGTGGATAAACTTATCTTGTCTGGACACGCCGCTCTTTCTCCTGG
AATCAGATGAAATGGAGAACTTGCTGACCTACAGCGGAGAGCCATAGAGCACGTGCTGCAGGTAGAGGCCTCCAGGAGCCC
TCGCACGTGTTTCAGCCTGAAGCAGCTGCTGCAGAGGTTACTGAAGAGCAATAGCCACTTGAGTGAGGAGTGCGGGGAGCTTCT
CCTGCAAAGAGGAACCCAGAGGTGGCCACAGGCTCTGGTTCTGAACAGAGACCAGAGGCTCGCCTGGGCAGAGAACAGCATTG
ACTTCATCAGCAGGAGCTGTGTGCGCATTCATCAGGAAGCTGCAGGCCCATGTCTGTTGATCAAAGCAGTCCACGGATAT
TTTGATTCAGACAGAATTACTCTGAGAAGGAGTCCCTGTCTGTTTCATGATAGACACGATGAAATCCACCCTCAAAGAGCAGTT
CCAGTTTGTGGAAGTCCAGGCAATCACTGTGTCCACATGAGCGAACCCAGCACGTGGCCAGTATCATCAGCTCCTTCTTAC
AGTGCACACACATGCTCCAGCCAGCTGTAG

```

**FIGURE 17C. Predicted amino acid sequence for the human homolog of CG3943 (315 amino acids), (SEQ ID NO:23)**

```

MSENAAPGLISELKLAVPWGHIAAKAWGSLQGPPVLCILHGWLNDASSFDRLLIPLLPQDFYYVAMDFGGHGLSSHYS
QTFVSEIRRVAALKWNRFSILGHSFSGGVVGMFFCTFPEMVDKLILDTPLFLLESDEMENLLTYKRRRAIEHVLQVEASQEP
SHVFSKQLLQRLKLSNSHLSEECGELLQQRGTTKVATGLVLNRDQLAWAENSIDFISRELCAHSIRKLQAHVLLIKAVHGY
FDSRQNYSEKESLSFMIDTMKSTLKEQFQFVEVPGNHCVMHSEPHVASIISFLQCTHMLPAQL.

```

drosophila	M G Q T R V A A T T A A Q S P A E H S P E T N G Q T E E P	30
mS0273353.1	H - - - - - A - - - - - G L H G L E - - - - -	6
HSC140179.1	L S E - - - - - N A A P C I S - - - - -	12
drosophila	L Q L L G E D S W E E F S I A Y E V G T Y E L K W A G S K E	60
mS0273353.1	- - - - - L K L A Y E V G H L A L K Y V G S Q K	25
HSC140179.1	- - - - - L L L A A Y P W G H L A A A A A M G S L Q	31
drosophila	R Q P I I A L H C M Q D N C G S R D E L C E L L E A T S I	90
mS0273353.1	N P E Y L G L H C M Q D N A N S R D E L L L L L L O D T C Y	55
HSC140179.1	G E P Y L G L H C M Q D N A S S R D E L L L L L L O D A Y Y	61
drosophila	L A I D L P G H G K S S H Y P M G M Q Y F I - - W D G I C	118
mS0273353.1	M A M D L G G H G L S S H Y N P G L P Y Y Q O N - - - Y S	82
HSC140179.1	V A M D L G G H G L S S H Y S P G Y P Y Y L O T E - - - Y S	88
drosophila	L T R R I M R K Y N M K N Y T T F G H S L G G A L T F M Y A	148
mS0273353.1	E Y R R V A T A F K W N Q E T L L G H S F G G C V G G T F A	112
HSC140179.1	E L R R V Y A A A L K W N R F S I L G H S F G G Y V G G M F F	118
drosophila	A S E P T E V E K L I N I D I A G E T Y R G T Q R M A E G T	178
mS0273353.1	G M E P E M Y D K L L L L D - S T P E F F D S N E M E N I T	141
HSC140179.1	C T E P E M Y D K L L L L D - T P L E L L E S D E A E N L T	147
drosophila	G - - - R A L D K F I D Y F T L P E S K Q P C Y S Y D E M I	205
mS0273353.1	T Y R R R N L E H T L O V E A S O K K S L R A Y S P E M E	171
HSC140179.1	T Y K R R A L E H Y L O V E A S O E P S - H Y F S L K Q L E	176
drosophila	K L V L D A Y D G S V D E P S V R V M N R G M R H N P S K	235
mS0273353.1	Q G F L N N - N S H L D K D C G E H I L O R G T - - T K V D	198
HSC140179.1	Q R L L K S - N S H L S E E C G E L L L L O R G T - - L K V A	203
drosophila	N G Y L F A P D L R L K - - V S L L G M F T A E Q T L A Y A	263
mS0273353.1	A G L V L N R D R R I S W P E N S F D E Y S K E M F V H S A	228
HSC140179.1	T G L V L N R D Q R L A W A E N S I D E I S R E L C A H S I	233
drosophila	R Q I R C R Y L N I R G I P G M K E E T P Q Y - Y A D - -	289
mS0273353.1	K S L O A S Y L M L K A L Q G Y - Y E V R R A N D A D K A P	257
HSC140179.1	R K L O A H V L L T K A V H G Y - E D - S R Q N Y S E K E S	261
drosophila	- - - Y L A T T R E N - A A K Y V Y A P Y P G T H H L H L Y	315
mS0273353.1	M H F M Y D T L R S T L K E R E O E Y E V E G N H Y I H M N	287
HSC140179.1	L S F M L D T M K S T L K E Q E O E Y E V E G N H C V H M S	291
drosophila	T P D R V A P H I I R L L K E - - - - - A	331
mS0273353.1	K P O V V A G V V G P E L O G L Q R M T S A R I .	312
HSC140179.1	E P O H V A S I T S S F L O C T H - M L P A Q L .	315

**FIGURE18. HUMAN HOMOLOG OF CG5216 (Sir2)****FIGURE 18A. BLASTN SEARCH RESULTS FOR CG5216**

Homology to human gene gi|14748197 sirtuin 1 ref|XP\_008902.2 REF-NP\_036370

/protein=DTP07482029.1 /gene=DTG07482004.1 /locus=DTL07482005.1  
 /garid=G2MS6DQFPDRQR /chrom=10 /contig=NT\_024033.3 /start=132574 /end=164444  
 /strand=plus  
 Similar to: gi|13630236|ref|XP\_008902.2| sirtuin (silent mating type information  
 regulation 2, *S.cerevisiae*, Length = 2244

Score = 410 bits (1042), Expect = e-114  
 Identities = 204/353 (57%), Positives = 256/353 (71%)  
 Frame = +1

Query: 153 WLQREFYTGRVPRQVIASIMPHFATGLAGDTDDSVLWDYLAHLNPKRRNKLASVNTFD 212  
 ++Q+ G PR ++ ++P T + DD LW + ++L+EP +R K +NT +  
 Sbjct: 559 FVQQHLMIGTDPRTILKDLLPE--TIPPELDDMTLWQIVINILSEPPKRRKKRKDINTIE 732

Query: 213 DVISLVKKSQKIIIVLTGAGVSVSCGIPDFRSTNGIYARLAHDFPDLDPQAMFDINIFYKR 272  
 D + L+++ +KIIIVLTGAGVSVSCGIPDFRS +GIYARLA DFPDLDPQAMFDI YF++  
 Sbjct: 733 DAVKLLQECKKIIIVLTGAGVSVSCGIPDFRSRDGIYARLAVDFPDLDPQAMFDIEYFRK 912

Query: 273 DPRPFYKFAREIYPGEFQPSPCRHRFIKMLETKGKLLRNYTQNIDTLERVAGIQRVIECHG 332  
 DPRPF+KFA+EIYPG+FQPS CH+FI + + +GKLLRNYTQNIDTLE+VAGIQR+I+CHG  
 Sbjct: 913 DPRPFKFAKEIYPGQFQPSLCHKFIALSDEKGLLRNYTQNIDTLEQVAGIQRRIQCHG 1092

Query: 333 SFSTASCTKCRFKCNADALRADIFAQRIPVCPQCQPNKEQSVDAVAVTEELRQLVENG 392  
 SF+TASC C++K + +A+R DIF Q +P CP+C +E L  
 Sbjct: 1093SFATASCLICKYKVDCEAVRGDIFNQVPRCP-----ADEPL-----A 1215

Query: 393 IMKPDIVFFGEGLPDEYHTVMATDKDVCDDLIVIGSSLKVRPVAHIPSSIPATVPQILIN 452  
 IMKP+IVFFGE LP+++H M DKD DLLIVIGSSLKVRPVA IPSSIP VPQILIN  
 Sbjct: 1216IMKPEIVFFGENLPEQFHRAMKYDKDEVDDLIVIGSSLKVRPVALIPSSIPHEVPQILIN 1395

Query: 453 REQLHHLKFDVELLGDSDVIINQICHRLSDNDDCWRQLCCDESVLTESKELMP 505  
 RE L HL FDVELLGD DVIIN++CHRL + +LCC+ L+E E P  
 Sbjct: 1396REPLPHLHFDVELLGDSDVIINELCHRLGGE---YAKLCCNPVKLSEITEKPP 1545

**FIGURE 18B. Predicted nucleotide sequence encoding the human homolog protein (2244 bp) (SEQ ID NO:24)**

>DTP07482020

ATGGCGGACGAGGCGGCCCTCGCCCTTCAGCCCGCGGCTCCCCCTCGGCGGCGGGGCGGACAGGGAGG  
 CCGCGTCTCCCGCGCGGGAGCCGCTCCGCAAGAGCGCGGAGAGATGGTCCCGGCTCGAGCGGAG  
 CCCGGGCGAGCCCGGTGGGGCGGCCCCAGAGCGTGAGGTGCCGCGCGGCCAGGGGCTGCCCGGGTGCG  
 GCGCGCGCGCGCTGTGCGCGGGAGGCGAGGCGAGGCGGCGCGGCGGCGGCGGAGCAAGAGGCCAGG  
 CGACTGCGGCGGCTGGGGAAGGAGACAATGGGCCGCGGCTGCAGGGCCCATCTCGGAGCCACCGCTGGC  
 CGACAACCTGTACGACGAAAGACGACGACGAGGCGGAGGAGGAGGAAGAGGCGGCGGCGGCGGCGGAT  
 GGGTACCGAGATAACCTTCTGTTCCGTTGATGAAATTATCACTAATGGTTTTCATTCTGTGAAAGTGATG  
 AGGAGGATAGAGCCTCACATGCAAGCTCTAGTACTGGACTCCAAGGCCACGGATAGGTCCATATACTTT  
 TGTTCAGCAACATCTTATGATTGGCACAGATCTTCGAACAATCTTAAAGATTTATGCGCGAAACAATA  
 CCTCCACCTGAGTTGGATGATATGACACTGTGGCAGATTGTTATTATATCTTTCAGAACCAACAAAAA  
 GGAAAAAAGAAAAATATTAATACAATTGAAGATGCTGTGAAATTAAGTCAAGAGTGCAAAAAAATAT  
 AGTTCTAACTGGAGCTGGGGTGTCTGTTTCATGTGGAATACCTGACTTCAGTCAAGGGATGGTATTTAT  
 GCTCGCCTTGCTGTAGACTTCCAGATCTTCAGATCTCAAGCGATGTTTGATATGAATATTTTCAGAA  
 AAGATCCAAGACCATCTTCTCAAGTTTGCAGAAAGGAAATATATCTGGACAATTCAGCCATCTCTCTGTCA  
 CAATTCATAGCCTTGTGATAGGAAAGGAACTACTTCGCAACTATACCCAGAACATAGACACGCTG  
 GAACAGGTTGCGGGAATCCAAGGATAATTCACTGTCATGGTTCTTTGCAACAGCATCTTGCCTGATTT

GTAAATACAAAGTTGACTGTGAAGCTGTACGAGGAGATATTTTTAATCAGGTAGTTCCTCGATGTCCTAG  
 GTGCCCAGCTGATGAACCGCTTGCTATCATGAAACCAGAGATTGTGTTTTTGGTGAAAATTTACCAGAA  
 CAGTTTCATAGAGCCATGAAGTATGACAAAGATGAAGTTGACCTCCTCATTTGTTATTGGGTCTTCCCTCA  
 AAGTAAGACCAGTAGCACATAATTCCAAGTTCCATACCCCATGAAGTGCCTCAGATATTAATTAATAGAGA  
 ACCTTTGCCCTCATCTGCATTTTGTATGTAGAGCTTCTTGGAGACTGTGATGTCATAATTAATGAATTGTGT  
 CATAGGTTAGGTGGTGAATATGCCAACTTTGCTGTAACCCTGTAAAGCTTTCAGAAATTAAGTAAAAAC  
 CTCCACGAACACAAAAAGAATTGGCTTATTTGTGACAGATTGCCACCCACACCTCTTCATGTTTCAGAAGA  
 CTCAAGTTCACCAGAAAGAACTTCACCACCAGATTCTTCAGTGATTGTACACTTTTAGACCAAGCAGCT  
 AAGAGTAATGATGATTTAGATGTGTCTGAATCAAAAGGTTGTATGGAAGAAAAACACAGGAAGTACAAA  
 CTTCTAGGAATGTTGAAAGTATTGCTGAACAGATGGAATAATCCGGATTGGAAGAAATGTTGGTTCTAGTAC  
 TGGGGAGAAAAATGAAAGAACTTCAGTGGCTGGAACAGTGAGAAAAATGCTGGCCTAATAGAGTGGCAAAG  
 GAGCAGATTAGTAGGCGGCTTGATGGTAATCAGTATCTGTTTTTGCACCAAATCGTTACATTTTCCATG  
 GCGCTGAGGTATATATCAGACTCTGAAGATGACGTCTTATCCTCTAGTTCTTGTGGCAGTAACAGTGATAG  
 TGGGACATGCCAGAGTCCAAGTTTAGAAGAACCCATGGAGGATGAAAGTGAAATGGAAGAAATCTACAAAT  
 GGCTTAGAAGATGAGCCTGATGTTCCAGAGAGAGCTGGAGGAGCTGGATTGTTGGGACTGATGGAGATGATC  
 AAGAGGCAATTAATGAAGCTATATCTGTGAAACAGGAAGTAACAGACATGAACATATCCATCAACAAATC  
 ATAG

**FIGURE 18C. Predicted amino acid sequence of the human homolog Protein (747 aa)  
 (SEQ ID NO:25)**

>DTP07482029

MADEAALALQPGGSFSAAGADREAASSPAGEPLRKRPRRDGPGLESPGEFPGAAPERIEVPAAARGCPGA  
 AAAALWREAEEAAAAGGEQEAQATAAAGEGDNGPGLQGPSREPPLADNLYDEDDDDDEGESEEEAAAAAI  
 GYRDNLLFGDEIITNGFHSCEDEEDRASHASSDWTFRPRIGPYTFVQQHLMIGTDPRTILKDLLPETI  
 PPELDDMTLWQIVINILSEPPKRKRKRDINTIEDAVKLLQECKKIIIVLTGAGVSVSCGIPDFRSRDGIY  
 ARLAVDFPDLDPQAMFDIEYFRKDPFPFFKFAKEIYPGQFQPSLCHKFIALSDKEGKLLRNYTQNIDTL  
 EQVAGIQRIIQCHGSFATASCLICKYKVDCEAVRGDIFNQVVPKPCPCPADEPLAIMKPEIVFFGENLPE  
 QFHRAMKYDKDEVDLLIVIGSSSLKVRPVALIPSSIPHEVPQILINREPLPHLHFDVELLGDVILINELC  
 HRLGGEYAKLCCNPVKLSEITEKPPRTQKELAYLSELPPPLHVSEDSSSPERTSPPDSSVIVTLLDQAA  
 KSNDDLVSSESKGCMEEKPQEVQTSRNVESIAEQMENPDLKNVGSSTGEKNERTSVAGTVRKCWPNRVAK  
 EQISRRLDGNQYLFLPPNRYIFHGAEVYDSEDDVLSSSSCGSNSDSGTCQSPSLEEPMEDESEIEEFYN  
 GLEDEPDVPERAGGAGFTGDGDDQEAINEAISVKQEVTDMMNYPNSKS\*

**FIGURE19. HUMAN HOMOLOG OF CG3758 (escargot)****FIGURE 19A. BLASTN SEARCH RESULT**

Homology to human gene ref|XP\_030528.1|

/protein=DTP06368020.1 /gene=DTG06368003.1 /locus=DTL06368005.1  
 /garid=G2KJ5GW5ZJC07Q /chrom=8 /contig=NT\_023679.3  
 /start=17601 /end=20059 /strand=plus Similar to:  
 gi|9187356|emb|CAB96946.1| (AL365370) hypothetical protein, similar to  
 (AB021644) GONADOTROPIN, Length = 807

Score = 240 bits (607), Expect = 2e-63  
 Identities = 109/137 (79%), Positives = 118/137 (85%)  
 Frame = +1

Query: 308 RYQCPDCQKSYSTFSGLTQHQQFHC PAAEGNQVKSFSCKDCDKTYVSLGALKMHIRTHT 367  
 ++QC C K+YSTFSGLT KH+Q HC A Q +KSFSCK CDK YVSLGALKMHIRTHT  
 Sbjct: 379 KFQCNLCNKTYSTFSGLAKHKQLHCDA----QSRKSFSCKYCDKEYVSLGALKMHIRTHT 546

Query: 368 LPCKCNLCGKAFSRPWLLQGHIRTHTGEEKPFSCQHCHRAFADRSNLRAHLQTHSDIKKYS 427  
 LPC C +CGKAFSRPWLLQGHIRTHTGEEKPFSC HC+RAFADRSNLRAHLQTHSD+KKY  
 Sbjct: 547 LPCVCKICGKAFSRPWLLQGHIRTHTGEEKPFSCPHCNRAFADRSNLRAHLQTHSDVKKYQ 726

Query: 428 CTSCSKTFSRMSLLTKH 444  
 C +CSKTFSRMSLL KH  
 Sbjct: 727 CKNCSTFSRMSLLHKH 777

Score = 31.3 bits (69), Expect = 2.5  
 Identities = 12/28 (42%), Positives = 18/28 (63%)  
 Frame = +1

Query: 308 RYQCPDCQKSYSTFSGLTQHQQFHC PAA 335  
 +YQC +C K++S S L KH++ C A  
 Sbjct: 718 KYQCKNCSTFSRMSLLHKHEESGCCVA 801

**FIGURE 19B. Predicted coding sequence for the human homolog protein (807 bp)  
(SEQ ID NO:26)**

>DTT06368011

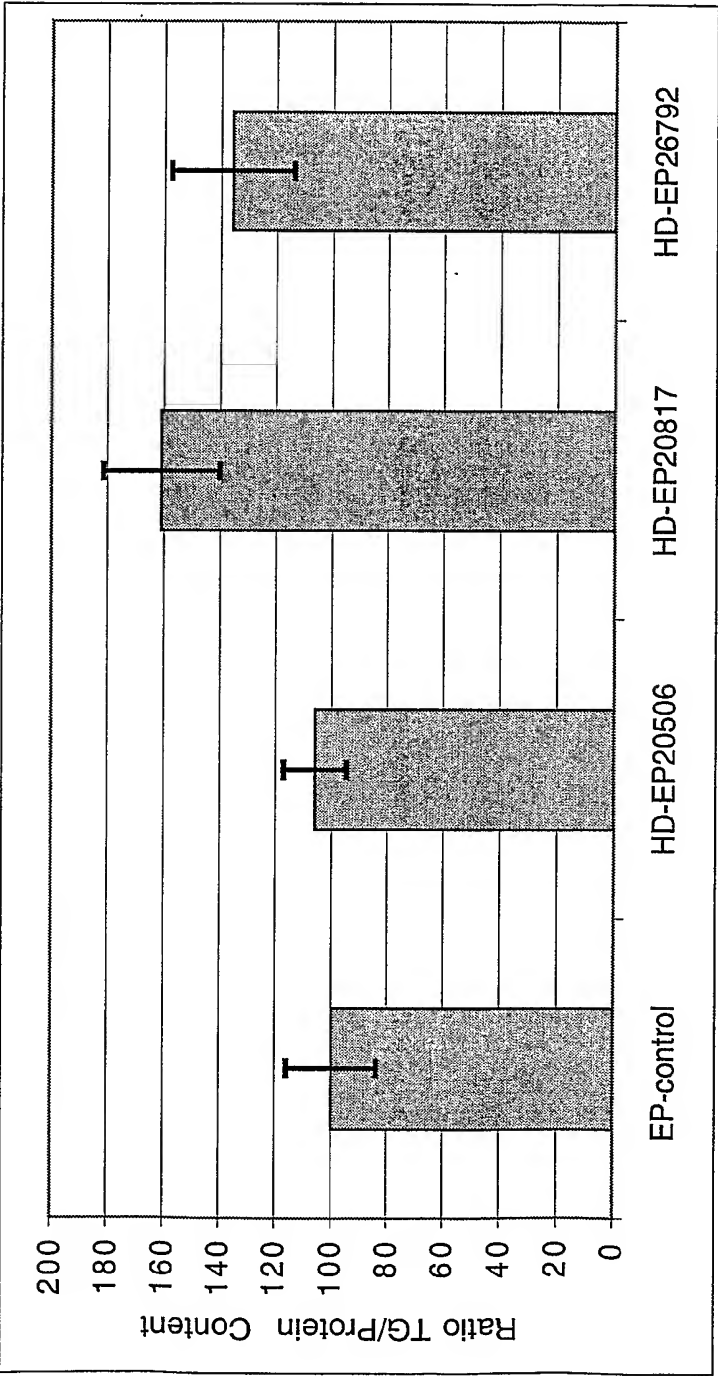
ATGCCGCGCTCCTTCCTGGTCAAGAAGCATTTCAACGCCTCCAAAAAGCCAAACTACAGCGAACTGGACA  
 CACATACAGTGATTATTTCCCCGTATCTCTATGAGAGTTACTCCATGCCTGTCATACCACAACCAGAGAT  
 CCTCAGCTCAGGAGCATACAGCCCCATCACTGTGTGGACTACCGCTGCTCCATGCCACGCCAGCTACCC  
 AATGGCCTCTCTCCTCTTTCCGGATACTCCTCATCTTTGGGGCGAGTGAGTCCCCCTCCTCCATCTGACA  
 CCTCCTCCAAGGACCACAGTGGCTCAGAAAGCCCCATTAGTGATGAAGAGGAAAGACTACAGTCCAAGCT  
 TTCAGACCCCCATGCCATTGAAGCTGAAAAGTTTTCAGTGCAATTTATGCAATAAGACCTATTCAACTTTT  
 TCTGGGCTGGCCAAACATAAGCAGCTGCACCTGCGATGCCAGTCTAGAAAATCTTTCAGCTGTAAATACT  
 GTGACAAGGAATATGTGAGCCTGGGCGCCCTGAAGATGCATATTCCGACCCACACATTACCTTGTGTTTG  
 CAAGATCTGCGGCAAGGCGTTTCCAGACCCTGGTTGCTTCAAGGACACATTAGAACTCACACGGGGGAG  
 AAGCCTTTTTCTTGCCCTCACTGCAACAGAGCATTTGCAGACAGGTCAAATCTGAGGGCTCATCTGCAGA  
 CCCATTCTGATGTAAAGAAATACCAGTGCAAAAACCTGCTCCAAAACCTTCTCCAGAAATGCTCTCTCTGCA  
 CAAACATGAGGAATCTGGCTGCTGTGTAGCACACTGA

**FIGURE 19C. Predicted amino acid sequence for the human homolog protein (268 aa)  
(SEQ ID NO:27)**

&gt;DTP06368020

MPRSFLVKKHFNASKKPNYSELDTHTVIISPVLYESYSMPVIPQPEILSSGAYSPITVWTTAAPFHAQLP  
NGLSPLSGYSSSLGRVSPPPPSDTSSKDHSGSESPISDEEERLQSKLSDPHAIEAEKFQC�LCNKTYSTF  
SGLAKHKQLHCDAQSRKSFSCKYCDKEYVSLGALKMHIRTHTLPCVCKICGKAFSRPWLLQGHIRTHTGE  
KPFSCPHCNRAFADRSNLRAHLQTHSDVKKYQCKNCSKTFSRMSLLHKHEESGCCVAH\*

FIGURE 20. Triglyceride content of Drosophila escargot (Gadfly Acc. No. CG3758) mutants



**FIGURE 21: THE HUMAN HOMOLOG OF CG3241 (msl-2)****FIGURE 21A: BLASTP SEARCH RESULTS FOR CG3241****Homology to human gene AB046805.1; BAB13411.1 Protein**

```

/protein=DTP02927030.1 /gene=DTG02927004.1 /locus=
  /garid=G2CHQPV_9KRKPS8 /chrom=3 /contig=NT_025665.2
  /start=189494 /end=208388 /strand=minus Similar to:
  gi|10047245|dbj|BAB13411.1| (AB046805) KIAA1585 protein
  [Homo sapiens]
  Length = 1734

```

Score = 58.2 bits (138), Expect = 3e-08, Identities = 24/66 (36%),  
Positives = 39/66 (58%), Frame = +1

```

Query: 50 DPYSPKGKRCQHNVCLRLGRGKHLFPSC+QCEGCSDFKTYEENRMAAQLLCYKTL CVH 109
      DP +P CQH VC+ C K + PSC+ C+ D++ +EEN+ ++ + CYK LC +
Sbjct: 157 DPIAPTNSTCQHYVCKTCKGKKMMMKPSCSWCK---DYEQFEENKQLSILVNCYKKLCEY 327

```

```

Query: 110 LLHSAL 115
      + + L
Sbjct: 328 ITQTTL 345

```

Score = 47.6 bits (111), Expect = 5e-05, Identities = 16/41 (39%),  
Positives = 25/41 (60%), Frame = +1

```

Query: 525 CRCGISGSSNTLTTCRNSRCPYKSYNSCAGCHCVCKNPH 565
      C+CG + + ++ TCR RCPY + +C C C C+N +
Sbjct: 1384 CKCGRATQNPVLTCTRGQRCPYSNRKACLDCICRGQCNSY 1506

```

**FIGURE 21B: Predicted coding sequence for the human homolog with Accession Number AB046805.1, encoding for the hypothetical KIAA1585 protein (1734 bp) (SEQ ID NO:28)**

>DTP02927021

```

ATGAACCCCGTGAATGCTACTGCTCTCTACATTTCCGCGAGCCGCCTAGTGCTCAACTACGACCCCGGAG
ACCCCAAGGCGTTTACTGAGATTAACAGGCTCTTGCCCTTACTTCCGACAGTCCCTTTTCGTGCTGTGTTTG
CGGACATTTGCTACAAGATCCTATTGCACCCACCAACTCCACCTGCCAACATTATGTCTGCAAACTTGT
AAAGGCAAGAAAATGATGATGAAACCTTCCCTGTAGCTGGTGCAAAGACTATGAGCAGTTTGAGGAAAACA
AGCAGTTAAGCATCCTAGTGAACCTGCTACAAAAAACTATGCGAGTATATAACACAGACTACACTGGCACG
GGATATAATAGAAGCAGTTGACTGTTCTTCTGATATTTTGGCTTTGCTTAATGATGGATCATTGTTTTGT
GAGGAGACAGAAAAACCCTCAGATTCATCCTTTACTTTGTGTTTGACACATTCCTTTTACCTTCAACCT
CAGAACCACAACTGATCCTCAAGCTAGTTTATCTCCAATGTCTGAAAGCACCTCAGCATTGCTATTGG
CAGTTCTGTTATCAATGGTTTGCCTACTTATAATGGGCTTTCAATAGATAGATTTGGTATAAAATATTCCT
TCACCTGAACATTCAAATACGATTGACGTATGTAATACTGTTGACATAAAAACTGAGGATCTGTCTGACA
GCCTGCCACCCGTTTGTGACACAGTAGCCACTGACTTATGTTCCACAGGCATTGATATCTGCAGTTTCAG
TGAAGATATAAAACCTGGAGACTCTCTGTTACTGAGTGTGAGGAAGTACTCCGCAGCTTAGAAAACCTGTT
TCAAATACAGAGGTCTGTTGCCCTAATTTGCAGCCGAACCTTGAAGCCACTGTATCCAATGGACCTTTTC
TGCAGCTTTCTTCCAGTCTCTTAGCCATAATGTTTTTATGTCCACCAGTCTTGCATTCATGGGTATC
ATGTACAGCAGCAACTCCGAAGATAGCAAAATTTGAATAGAAAACGATCCAGATCAGAGAGTGACAGTGAG
AAAGTTCAGCCACTTCCAATTTCTACCATTTATCCGAGGCCCAACACTGGGGGCATCTGCTCCTGTGACAG
TGAAACGGGAGAGCAAAATTTCTCTTCAACCTATAGCAACTGTTCCCAATGGAGGCACAACACCTAAAAT
CAGCAAACTGTACTTTTATCTACTAAAAGCATGAAAAAGAGTCATGAACATGGATCCAAGAAATCTCAC

```

TCTAAAACCAAGCCAGGTATTCTTAAAAAAGACAAAGCAGTAAAGGAAAAGATTCTTAGTCATCATTTTA  
TGCCAGGAAGTCCTACCAAGACTGTGTACAAAAAACCCAGGAAAAGAAAGGGTGTAATGTGGGCGTGC  
TACTCAAAATCCAAGTGTCTTACATGCCGAGGCCAACGCTGCCCTTGCTACTCTAACCGCAAAGCCTGC  
TTAGATTGTATATGTCGTGGCTGCCAAACTCCTATATGGCCAATGGGGAGAAGAAGCTGGAGGCATTTG  
CCGTGCCAGAAAAGGCCTTGGAGCAGACCAGGCTCACTTTGGGCATTAACTGACTAGCATTGCTGTGCG  
TAACGCTAGTACCAGCACCAGTGTATAAATGTCACAGGGTCCCCAGTAACGACGTTTTTAGCTGCCAGT  
ACACATGATGATAAAAGTTTGGATGAAGCTATAGACATGAGATTGACTGTAA

**FIGURE 21C: Predicted amino acid sequence for the human homolog of CG3241 (577 aa), (SEQ ID NO:29)**

>DTP02927030

MNPVNATALYISASRLVLNYDPGDPKAFTEINRLLPYFRQSLSCCVCGHLLQDPIAPTNSTCQHYVCKTC  
KGKMMMKPSCSWCKDYEQFEENKQLSILVNCYKKLCEYITQTTLARDIIEAVDCSSDILALLNDGSLFC  
EETEKPSDSSFTLCLTHSPLPSTSEPTTDPQASLSPMSESTLSIAIGSSVINGLPTYNGLSIDRFGINIP  
SPEHSNTIDVCNTVDIKTEDLSDSLPFVCDTVATDLCSTGIDICSFSEDIKPGDSLSSVEEVLRSLTV  
SNTEVCCPNLQPNLEATVSNPFLQLSSQSLSHNVFMSTSPALHGLSCTAATPKIAKLNRKRSRSESDSE  
KVQPLPISTIIRGPTLGASAPVTVKRESKISLQPIATVPNGGTPKISKTVLLSTKSMKKSHEHGSKKSH  
SKTKPGILKKDKAVKEKIPSHHFMPGSPTKTVYKKPQEKKGCKCGRATQNPSVLTCTRGQRCPCYSNRKAC  
LDCICRGQNSYMANGEKKLEAFVPEKALEQTRLTLGINVTSIAVRNASTSTSVINVTGSPVTTFLAAS  
THDDKSLDEAIDMRFDC

**FIGURE 21D: ClustaW alignment of *Drosophila* msl-2 (GadFly Acc. No. CG 3241; referred to as 'd Msl-2') and human (hHIA1585) and mouse (mBF471233) homologs. The sequences are shown in the one letter code; shaded residues match exactly.**

```

d Msl-2  MAQTA  --- YLKVTRIAMRSA SMLSKRRV EELNSGLGELR
hHIA1585  MNPVNATALVISASRLVLNYDPG-DPKAFTEIARLLPYFR
mBF471233  M-----

d Msl-2  QLLSCAVVCCQILVDPVSPKQKRCQHNVCRLCLRGGKHLER
hHIA1585  QSLSCAVVCGHLLQDEPIAETNSTCQH YVCKTQ-KGKKMMMK
mBF471233  Q-----

d Msl-2  SCTQCGCGCDENLYEENRMMAAQLLGYKTLGVHLLHSALE
hHIA1585  PS--GSWCKDYEQFEENKQLSILVNCYKKKLCBYITQTTLA
mBF471233  P-----

d Msl-2  GPPLAGMRPQVARRAVRRLKLPKATQERELREGSNISDT
hHIA1585  RDITFA-VDCSSDILALL-----NDGSLCEETETKPSDSS
mBF471233  R-----

d Msl-2  DIELPQPDLPFLKDMPTSLPABTPPTSAVTTPEIRYDHL
hHIA1585  TLCNTHSPDR-----STSEETTDHQAQLSPMSESTLSIAI
mBF471233  D-----

d Msl-2  NISDIEERAAALMAHQGRFSEFIRLFRGSRMGMLSHAGQIV
hHIA1585  GSSVINGLPT-----YNGLSIDRFGINIPSPESHSNTLD
mBF471233  N-----

d Msl-2  IATSSSESGEIDQAWATDQVDELSSGVSYSKMINSNNFAV
hHIA1585  V---CNTVDIKTEDLSDSLPPVCDTVATDLCSTGIDIC--
mBF471233  I-----

d Msl-2  YVMEPTSAATTKFDORHQQGQVYQMDSTQLAVEAAVEETV
hHIA1585  F-----SEDIKPIGDSLLLSVEEVRSLETYSNTE
mBF471233  Y-----

d Msl-2  ETSSTQILITLSSAARFVYETSTQDEYVLSGSESPRISQD
hHIA1585  VCCPNLQ---PNLEAATVSNQPFLLQLSSQSLSHNVFMSTSP
mBF471233  E-----

d Msl-2  NLOVMEESDBALVREEMVEEAEGTSTPGEVVAEEMMEDQHL
hHIA1585  AHHGLSCTA--TPKIAKLNRRKRSRSESDSEKV---QRP
mBF471233  N-----

d Msl-2  VHTSQSRTQTMBEAVSEHVAATKTOIGHVQTELDQDAESLQ
hHIA1585  ISHTIIRGPTLGASAPVTVKRESKISQPIAT-VPNGGTT
mBF471233  V-----

d Msl-2  RDPEDAKAAAEAKKKKKDPAHLSARFQKEPSDEPTDKRK
hHIA1585  K---ISATVLLSTSMKKS-----HEHGSKKSHSKT
mBF471233  R-----KKS-----HEHGSKKSHSKS

d Msl-2  RTRITLXASQALKIEPVPSVEVKTQVQSGGALRRIRRGIDKE
hHIA1585  KPGILKKKDKAVK-EKIPSHHFMFGSPTTVYKKPQEKKG-
mBF471233  KPGILKKKDKAVK-EKIPSHHFMFGSPTTVYKKPQEKKG-

d Msl-2  EVKQKPKPKKCKCQCGISGSSMTLTCTCNKRCPCYKANSQAG
hHIA1585  -----CKCGRATQNPSVLTCTGQRCPCYSNRKAQLD
mBF471233  -----CKCGRATQNPSVLTCTGQRCPCYSNRKAQLD

d Msl-2  CHCVCCKNHEHREDVSESDQDLEDFRMPKDVPEPMTQSE
hHIA1585  GICRGQONS---M-ANGEKKLEAFAVE-----
mBF471233  GICRGQONS---M-ANGEEEAGGICGAR-----

d Msl-2  PPVVAEPEOPEENSMAPEPDSAPISLVPLNNDDQSQRPHVL
hHIA1585  -----
mBF471233  -----

d Msl-2  VQNEKGEYQGENEFGSGKPLDPLVVGERTVQQLQHTDGLG
hHIA1585  -----KALEQTRLTLGENTV-----
mBF471233  -----KGLEQTRLTLGENTV-----

d Msl-2  STROYAVIMPTIIEENPEASLSPEPPEAPDREVIEDPPAK
hHIA1585  -----TSIAVRNASTSTSVINVTGSEVT
mBF471233  -----TSIAVRNASTSTSVINVTGSEVT

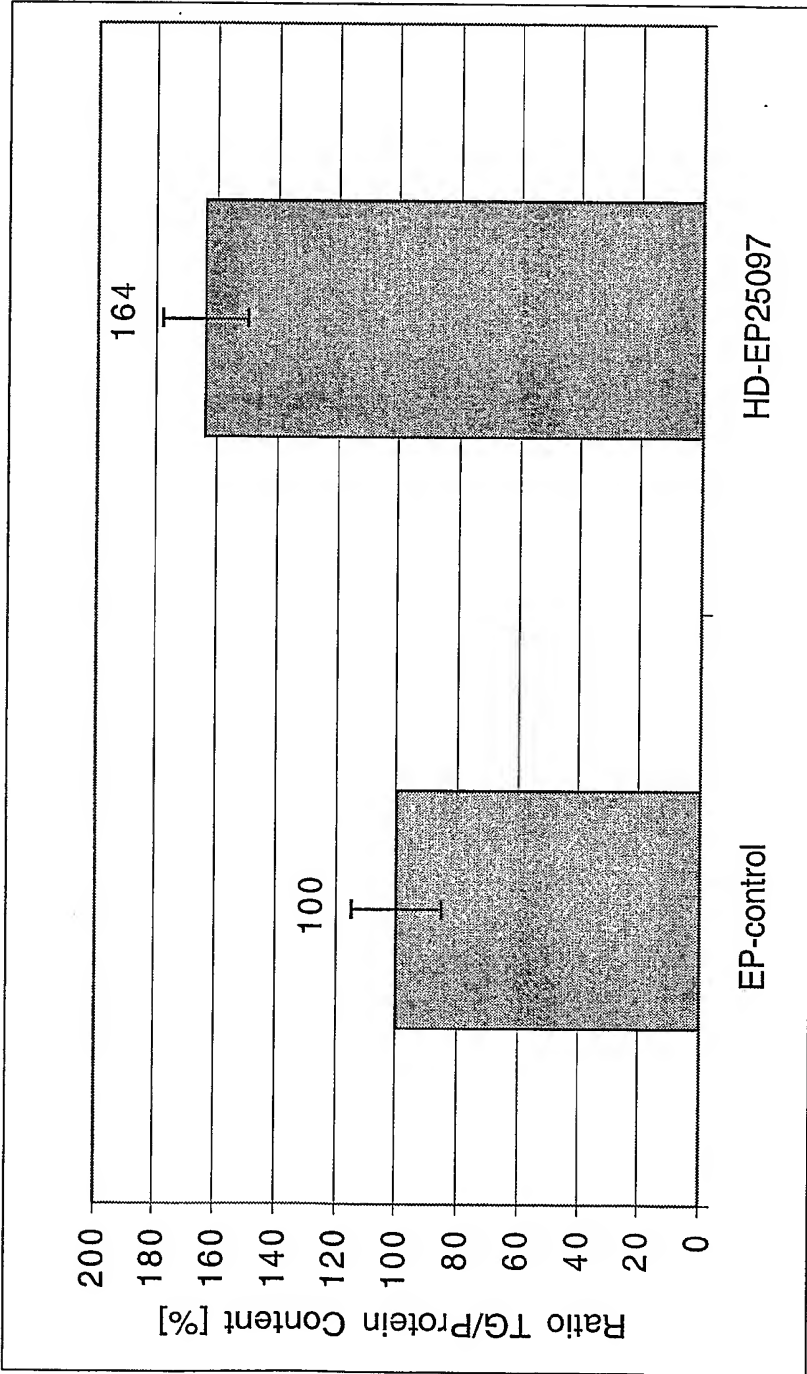
d Msl-2  KPRTRSRTRRGPAPEQAIDTVDELVSQGSRSNSAACDRSSA
hHIA1585  TFLAASTHDDKS-----
mBF471233  TFLAASTHDDKK-----

d Msl-2  TQNAHSLFEETIMSQSDDL
hHIA1585  -----LDEAIDMRFDIC
mBF471233  -----K K K K K A

```

Decoration 'Decoration #1': Shade (with solid bright yellow) residues that match d Msl-2 exactly.

FIGURE 22. Triglyceride content of a *Drosophila msl-2* (Gadfly Acc. No. CG3241) mutant



# **FIGURE 23. THE HUMAN HOMOLOG OF CG11940 and TG content of a Drosophila CG11940 mutant**

## **FIGURE 23A. BLASTN SEARCH RESULT FOR CG11940**

Homology to human gene : ref|XP\_028059.1| (XM\_028059) KIAA1681 protein [H. sapiens] BAB69020 Human alisin aslcr9 in database

/protein=DTP07670032.1 /gene=DTG07670002.1 /locus=DTL07670006.1  
 /garid=G2MX526\_9QLTJML /chrom=10 /contig=NT\_026379.1  
 /start=367384 /end=437764 /strand=minus Similar to:  
 gi|9506857|ref|NP\_061916.1| similar to proline-rich  
 protein 48 [Homo sapiens]  
 Length = 1497

Score = 198 bits (498), Expect = 3e-50  
 Identities = 109/266 (40%), Positives = 165/266 (61%), Gaps = 8/266 (3%)  
 Frame = +1

Query: 349 KADKIQLALHKLESAPIRRLFKAFSTSDGASKSLVDERMGCCHVTRLLADKNHVQMOSN 408  
 KADKI+LAL KL+ A +++L VK +D ++KSL+VDER V L +K H +  
 Sbjct: 475 KADKIKLALAKLKEAKVKLVVKVHMNDNSTKSLMVDERQLARDVLDNLFKETHCDCNVD 654

Query: 409 WALVEHLGDLQMERLFEDHELLVDNLMTWHS-DAGNRVLFQQRDPKVTFLRPE-LYLP 466  
 W L E +LQ+ER FEDHE +V+ L + D N++LF ++ +K +F P+ YL  
 Sbjct: 655 WCLYEIYPQLIERFFEDHENVVEVLSPDGTRDTENKILFLEKEEKYAVFKNPQNFYLDN 834

Query: 467 PQMAPGCQHDEQ---TRQMLLDEFDSHNQL--QMDGPLYMKADPKKGWKRYHFVLRSS 520  
 + +E+ ++ LL+E F + + +++G LY+K D KK WKR +F+LR+S  
 Sbjct: 835 RGKKEKETNEKMNKKNESLEESFCGTSTIVPELEGALYLKEDGKKSWKRRYFLLRAS 1014

Query: 521 GLYYFPKEKTKNTRDLACLNLFGHNVTGLGWRKKWKSPTDYTFGFKAVGDSSLGKSCR 580  
 G+YY PK KTK +RDLAC F N+Y G + K+K+PTDY F K + K +  
 Sbjct: 1015GIYYVPKGTKTSRDLACFIQFENVNIYYGTQHKMKYKAPTDYCFVLK---HPQIQKESQ 1185

Query: 581 SLKMLCAEDLPTLDRLWLTAIRVCKYKQLWDSHK 614  
 +K LC +D TL++W+ IR+ KYGK L+D+++  
 Sbjct: 1185YIKYLCDDTRTLNQWVMGIRIAKYGKTLYDNYQ 1287

## **FIGURE 23B. Predicted coding sequence for the human homolog of CG11940 (1497 bp) (SEQ ID NO:30)**

>DTT07670023

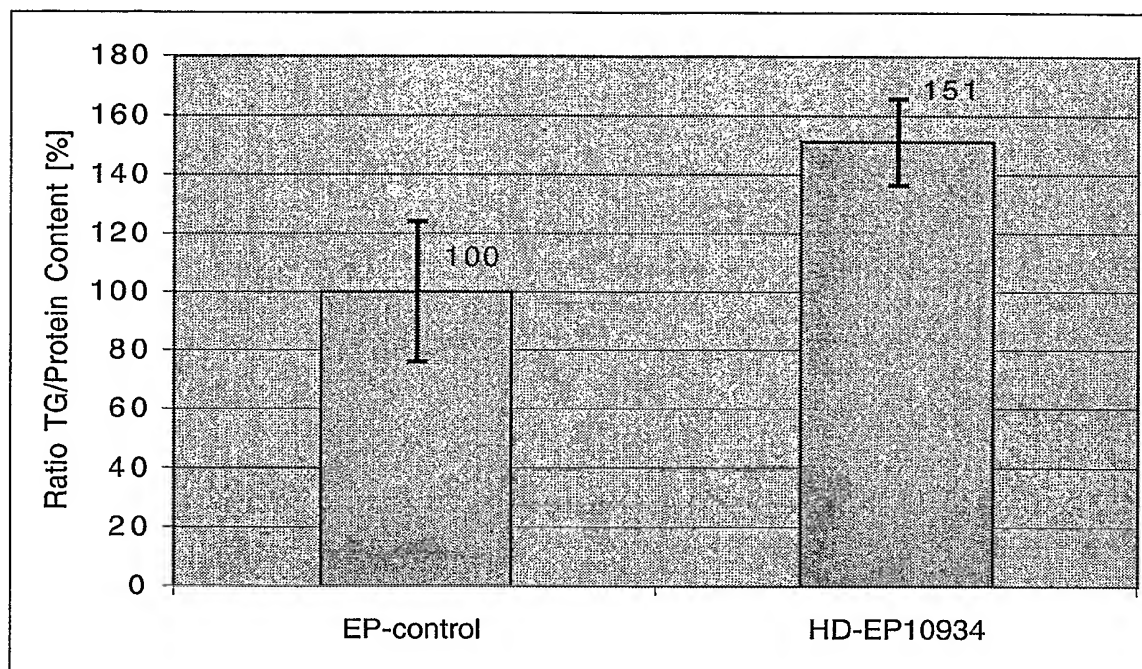
ATGGGTGAGTCAAGTGAAGACATAGACCAAATGTTTCAGCACTTTGCTGGGAGAGATGGATCTTCTGACTC  
 AGAGTTTAGGAGTTGACACTCTCCCTCCTCCTGACCTAATCCACCCAGAGCTGAATTTAACTACAGTGT  
 GGGGTTTAAAGATTTAAATGAGTCCTTAAATGCACTGGAAGACCAAGATTTAGATGCTCTCATGGCAGAT  
 CTGGTAGCAGACATAAGTGAGGCTGAGCAGAGGACAATCCAGGCACAGAAAAGAGTCCTTGCAGAATCAAC  
 ATCATTTCAGCATCTCTACAAGCATCAATTTTCAGTGGTGCAGCCTCTCTTGGTTATGGAACAAATGTTGC  
 TGCCACTGGTATCAGCCAATATGAGGATGACTTACCACCTCCACCAGCCGATCCTGTGTTAGACCTTCCA  
 CTGCCACCACCACCTCCTGAACCTCTCTCTCAGGAAGAGGAAGAAGCCCAAGCCAAGGCTGATAAAATTA  
 AGCTGGCGCTGGAAAACTGAAGGAGGCCAAGGTTAAGAAGCTCGTCGTCAAGGTGCACATGAATGATAA  
 CAGCACAAAGTCACATGATGGTGGATGAGCGACAGCTGGCCCGAGATGTTCTGGACAACCTTTTCGAGAAA  
 ACTCATTGTGACTGCAATGTAGACTGGTGTCTTTATGAAATCTACCCGGAACCTACAAATTGAGAGGTTTT  
 TTGAAGACCATGAAAATGTTGTTGAAGTCTTATCACCAGACGGGACAAGAGACACAGAAAATAAAATACT  
 ATTTTGTGGAGAAAGAGGAGAAATATGCTGTATTTAAAAACCCCAAGATTTCTACTTGGATAACAGAGGA  
 AAAAAAGAAAGCAAGGAAACTAATGAGAAAATGAATGCTAAGAACAAGGAATCCTTACTTGAGGAAAGTT

TCTGTGGAACATCTATCATTGTACCAGAACTGGAAGGAGCTCTTTATTTGAAAGAAGATGGAAAGAAATC  
CTGGAAAAGGCGCTATTTTCTTTTACGGGCTTCTGGAATTTATTATGTACCCAAAGGAAAGACTAAGACA  
TCTCGAGATCTGGCGTGTTTTATACAGTTTGAAAATGTCAACATTTACTATGGGACTCAGCATAAAATGA  
AATATAAAGCGCCCACTGACTATTGCTTTGTTTTAAAGCACCCCAAATTCAGAAGGAGTCCCAGTATAT  
CAAGTATCTCTGCTGTGATGACACAAGAACCCTTAACCAGTGGGTCATGGGAATACGGATAGCCAAGTAT  
GGGAAGACTCTCTATGATAACTACCAGCGGGCTGTGGCAAAGGCTGGACTTGCCTCTCGGTGGACAAACT  
TGGGGACAGTCAATGCAGCTGCACCAGCTCAGCCATCTACAGGACCTAAAACAGGCACCACCCAGCCCAA  
TGGACAGATTCCCCAGGCTACACATTCTGTCTCAGTGCTGTTCTCCAAGAGGCCAGAGACATGCTGAAACA  
TCGAAGGTAAAACCAGCAAGCAGCTGA

**FIGURE 23C Predicted amino acid sequence for the human homolog of CG11940  
(498 aa) (SEQ ID NO:31)**

>DTP07670032

MGESSEDIDQMFSTLLGEMDLLTQSLGVDLTPPPDPNPPRAEFNYSVGFKDLNESLNALEDQDLDALMAD  
LVADISEAEQRTIQAQKESLQNQHHSASLQASIFSGAASLGYGTNVAATGISQYEDDLPPPPADPVLDP  
LPPPPPEPLSQEEEEAQAKADKIKLLEKLKEAKVKLVVKVHMNDNSTKSLMVDRLARDVLDNLFK  
THCDCNVDWCLYEIYPELQIERFFEDHENVVEVLSPDGTRDTENKILFLEKEEKYAVFKNPQNFYLDNRG  
KKESKETNEKMNAKNKESLLEESFCGTSIIIVPELEGALYLKEDGKKS WKRRYFLLRASGIYYVPKGKTKT  
SRDLACFIQFENVNIYYGTQHKMKYKAPTDYCFVLKHPQIQKESQYIKYLCCDDTRTLNQWVMGIRIAY  
GKTLTDNYQRAVAKAGLASRWTNLGTVNAAAPAPSTGPKTGTTQPNGQIPQATHSVSAVLQEAQRHAET  
SKVKPASS\*

**FIGURE 23D. Triglyceride content of a *Drosophila* CG11940 (Gadfly Acc. No.) mutant**

**FIGURE 24. HUMAN HOMOLOG OF CG1624 (dappled)****FIGURE 24A. tBLASTN SEARCH RESULT FOR CG1624**

Homology to human gene ref XM\_067369; protein ref XP\_067369.1

/garid=G73KJ99\_DQ8KMK /chrom=3 /contig=NT\_028139.1 /start=1112 /end=7855  
/strand=plus

Score = 193 bits (485), Expect = 3e-48, Identities = 91/171 (53%),  
Positives = 118/171 (68%), Frame = +2

Query: 525 LSFATEGHEDGQVSRPWGLCVDKMGHVLVSDRRNNRVQVFNPDSGLKFKFGRKGVGNGEF 584  
LSF +EG DG++ RPWG+ VDK G+++V+DR NNR+QVF P G+ KFG G G+F  
Sbjct: 878 LSFSGEGSDGKLCRPWGVSVDEKGYIIVADRSNNRIQVFKPCGAFHHKFGTGLSRPGQF 1057

Query: 585 DLPAGICVDVDNRRIIVVDKDNHRVQIFTASGVFLKFGSYGKEYGQFQYPWDVAVNSRRQ 644  
D PAG+ D RI+V DKDNHR+QIFT G FLLKFG G + GQF YPWDVAVNS +  
Sbjct: 1058 DRPAGVACDASRRIVVADKDNHRIQIFTFEGQFLLKFGKGTNGQFNYPWDVAVNSEGK 1237

Query: 645 IVVTDSRNHRIQQFDSEGRFIRQIVFDNHGQTKGIASPRGVCYTPTGNIIV 695  
I+V+D+RNHRIQ F +G F+ + F+ K SPRGV + G+++V  
Sbjct: 1238 ILVSDTRNHRIQLFGPDGVFLNKYGFEG-ALWKHFDSPRGVAFNHEGHLVV 1387

Score = 122 bits (302), Expect = 9e-27, Identities = 66/171 (38%),  
Positives = 96/171 (55%), Gaps = 1/171 (0%), Frame = +2

Query: 525 LSFATEGHEDGQVSRPWGLCVDKMGHVLVSDRRNNRVQVFNPDSGLKFKFGRKGVGNGEF 584  
L F +G ++GQ + PW + V+ G +LVSD RN+R+Q+F PDG K+G +G F  
Sbjct: 1160 LKFGEKGTNGQFNYPWDVAVNSEGKILVSDTRNHRIQLFGPDGVFLNKYGFEGALWKHF 1339

Query: 585 DLPAGICVDVDNRRIIVVDKDNHRVQIFTASGVFLKFGSYGKEYGQFQYPWDVAVNSRRQ 644  
D P G+ + + ++V D +NHR+ + GS G GQF P VAV+ +  
Sbjct: 1340 DSPRGVAFNHEGHLVVTDFFNNHRLLVHPDCQSARFLGSEGTGNGQFLRPQGVAVDQEGR 1519

Query: 645 IVVTDSRNHRIQQFDSEGRFIRQIVFDNHGQTKG-IASPRGVCYTPTGNIIV 695  
I+V DSRNHR+Q F+S G F+ + F G G + P G+ TP G I+V  
Sbjct: 1520 IIVADSRNHRVQMFESNGSFLCK--FGAQSGFGQMDRPSGIAITPDGMIVV 1669

Score = 82.7 bits (201), Expect = 6e-15, Identities = 38/83 (45%),  
Positives = 56/83 (66%), Frame = +2

Query: 529 TEGHEDGQVSRPWGLCVDKMGHVLVSDRRNNRVQVFNPDSGLKFKFGRKGVGNGEFDLPA 588  
+EG +GQ RP G+ VD+ G ++V+D RN+RVQ+F +GS KFG +G G G+ D P+  
Sbjct: 1454 SEG TGNGQFLRPQGVAVDQEGRIIVADSRNHRVQMFESNGSFLCKFGAQSGFGQMDRPS 1633

Query: 589 GICVDVDNRRIIVVDKDNHRVQIF 611  
GI + D I+VVD N+R+ +F  
Sbjct: 1634 GIAITPDGMIVVVDGNNRILVF 1702

**FIGURE 24B. Predicted coding sequence for the human homolog of CG1624 (1545 base pairs), (SEQ ID NO:32)**

ATGAAGGCGAAGGTTGTGCAGTCGGAGGTCAAAGCCGTGACGGCGAGGCATAAGAAAGCCCTGGAGGAACGCGAGTGTGA  
GCTGCTGTGGAAGGTAGAAAAGATCCGCCAGGTGAAAGCCAAGTCTCTGTACCTGCAGGTGGAGAAGCTGCCGCCAAAACC  
TCAACAAGCTTGAGAGCACCATCAGTGCCTGTCAGCAGGTCTTGGAGGAGGGTAGAGCGCTAGACATCCTACTGGCCCGA  
GACCGGATGCTGGCCCGAGGTGCAGGAGCTGAAGACCGTGCAGGACCTCCTGCAGCCCCAGGAAGACGACCGAGTCATGTT  
CACACCCCCCGATCAGGCACTGTACCTTGCCATCAAGTCTTTTGGCTTTGTTAGCAGCGGGGCCCTTTGCCCCACTCACCA  
AGGCCACAGGCGATGGCCTCAAGCGTGCCCTCCAGGGTAAGGTGGCCTCCTTCACAGTCATTGGTTATGACCACGATGGT  
GAGCCCCGCCTCTCAGGAGGCGACCTGATGTCGGCTGTGGTCTTGGGCCCTGATGGCAACCTGTTTGGTGCAGAGGTGAG  
TGATCAGCAGAATGGGACATACGTGGTGAGTTACCGACCCAGCTGGAGGGTGAGCACCTGGTATCTGTGACACTGTGCA  
ACCAGCACATTGAGAACAGCCCTTTCAAGGTGGTGGTCAAGTCAGGCCGAGCTACGTGGGCATTGGGCTCCCGGGCCTG  
AGCTTCGGCAGTGAGGGTGACAGCGATGGCAAGCTCTGCCGCCCTTGGGGTGTGAGTGTAGACAAGGAGGGCTACATCAT  
TGTCGCCGACCGCAGCAACAACCGCATCCAGGTGTTCAAGCCCTGCGGCGCCTTCCACCACAAATTCGGCACCCCTGGGCT  
CCCGGCCTGGGCAGTTTCGACCGACCGCCGCGCTGGCCTGTGACGCCTCACGAGGATCGTGGTGGCTGACAAGGACAAT  
CATCGCATCCAGATCTTCACGTTTCGAGGGCCAGTTCCTCCTCAAGTTTGGTGAGAAAGGAACCAAGAATGGGCAGTTCAA  
CTACCCCTGGGATGTGGCGGTGAATTCTGAGGGCAAGATCCTGGTCTCAGACACGAGGAACCAACCGGATCCAGCTGTTTG  
GGCTGATGGTGTCTTCTTAAACAAGTATGGCTTCGAGGGGGCTCTCTGGAAGCACTTTGACTCCCCACGGGGTGTGGCC  
TTCAACCATGAGGGCCACTTGGTGTGCTACTGACTTCAACAACCAACCGGCTCCTGGTTATTACCCCCGACTGCCAGTCGGC  
ACGCTTTCTGGGCTCGGAGGGCACAGGCAATGGGCAGTTCCTGCGCCCAAGGGGTAGCTGTGGACCAGGAAGGGCGCA  
TCATTGTGGCGGATTCAGGAACCATCGGGTACAGATGTTTGAATCCAACGGCAGCTTCCTGTGCAAGTTTGGTGTCTCAA  
GGCAGCGGCTTTGGGCAGATGGACCGCCCTTCCGGCATCGCCATCACCCCCGACGGAATGATCGTTGTGGTGGACTTTTG  
CAACAATCGAATCCTCGTCTTCTAA

**FIGURE 24C. Predicted amino acid sequence for the human homolog of CG1624 (515 amino acids), (SEQ ID NO:33)**

MKAKVVQSEVKAVTARHKKALEERECELLWKVEKIRQVKAKSLYLQVEKLRQNLNKLESTISAVQQVLEEGRALDILLAR  
DRMLAQVQELKTVRSLLQPDQEDDRVMFTPPDQALYLAIKSFGFVSSGAFAPLTKATGDGLKRALQGKVASFTVIGYDHDG  
EPRLSGGDLMSAVVLGPDGNLFGAEVSDQQNGTYVVSYPQLGEHLVSVTLNQHIENSPFKVVVKSGRSYVVGIGLPGL  
SFGSEGDSDGKLCRPWGVSVDEKEYIIVADRSNNRIQVFKPCGAFHHKFGTGLSRPGQFDRPAGVACDASRRIVVADKDN  
HRIQIFTFEGQFLLKFGEKGTNGQFNYPWDVAVNSEGKILVSDTRNHRIQLFGPDGVFLNKYGFEGALWKHFDSPRGVA  
FNHEGHLVVTDFFNNHRLVVIHPDCQSARFLGSEGTNGQFLRPQGVAVDQEGRIIVADSRNHRVQMFESNGSFLCKFGAQ  
GSGFGQMDRPSGLAITPDGMIVVVDGNNRILVF

**FIGURE 25. HUMAN HOMOLOG OF CG11753****FIGURE 25A. tBLASTN SEARCH RESULT FOR CG11753**

Homology to human gene with GenBank Accession Number DTG11814022.1; protein with GenBank Accession Number DTP11814022.1

/protein=DTP11814022.1 /gene=DTG11814022.1 /locus=DTL11814022.1 /garid=G4RRSRH\_20MGXD  
 /chrom=20 /contig=NT\_011362.5 /start=9045086 /end=9049619 /strand=plus  
 Similar to: gi|14771151|ref|XP\_029849.1| similar to RIKEN cDNA 2610042O14 gene (M. musculus) [Homo sapiens], Length = 492

Score = 128 bits (318), Expect = 5e-30, Identities = 61/144 (42%),  
 Positives = 89/144 (61%), Frame = +1

Query: 4 GTFRNTQWDPTLLSSQIVSMQFCVYFTLGLLVFVANKLSGDNYSLDHLFEYHEIHIYDMG 63  
 G FR+ WDP L+ SQIV MQ Y +LGL + + + L + SLD +F+ +  
 Sbjct: 7 GQFRSYVWDPLLILSQIVLMQTVYYGSLGLWLALVDGLVRSSPSLDQMFDAILGFSTPP 186

Query: 64 GRLVICAFLVNAFLASLALWCIVRRAKLCLDFSCTFHVLHLLICWWYNRSFPANASWLL 123  
 GRL + +F+LNA +L L +RR K CLDF+ T H HLL CW+Y+ FP+ +WWL+  
 Sbjct: 187 GRLSMMSFILNALTALGLLYFIRRGKQCLDFTVTVHFFHLLGCWFYSSRFPALTWWLV 366

Query: 124 NVITGTMICIGGEFLCLQTEMKEI 147  
 + +M + GE+LC++TE+KEI  
 Sbjct: 367 QAVCIALMAVIGEYLCMRTELKEI 438

**FIGURE 25B. Predicted coding sequence for the human homolog of CG11753 (492 base pairs), (SEQ ID NO:34)**

ATGGCGGGTCAGTTCCGCAGCTACGTGTGGGACCCGCTGCTGATCCTGTGCGCAGATCGTCCTCATGCAGA  
 CCGTGTATTACGGCTCGCTGGGCCTGTGGCTGGCGCTGGTGGACGGGCTAGTGCGAAGCAGCCCCCTCGCT  
 GGACCAGATGTTTCGACGCCGAGATCCTGGGCTTTTCCACCCCTCCAGGCCGGCTCTCCATGATGTCCTTC  
 ATCCTCAACGCCCTCACCTGTGCCCTGGGCTTGCTGTACTTCATCCGGCGAGGAAAGCAGTGTCTGGATT  
 TCACTGTCACTGTCCATTTCTTTACCTCCTGGGCTGCTGGTTCTACAGCTCCCGTTTCCCCTCGGCGCT  
 GACCTGGTGGCTGGTCCAAGCCGTGTGCATTGCACTCATGGCTGTCATCGGGGAGTACCTGTGCATGCGG  
 ACGGAGCTCAAGGAGATAGGAGATAGGAATTTGCTGCTAAGATTTTCTTTGGGGTGGAGTTTCTCTGT  
 GA

**FIGURE 25C. Predicted amino acid sequence for the human homolog of CG11753 (163 amino acids), (SEQ ID NO:35)**

MAGQFRSYVWDPLLILSQIVLMQTVYYGSLGLWLALVDGLVRSSPSLDQMFDAILGFSTPPGRLSMMSF  
 ILNALTALGLLYFIRRGKQCLDFTVTVHFFHLLGCWFYSSRFPALTWWLVQAVCIALMAVIGEYLCMR  
 TELKEIGDRNLLLRFFFGVEFPL

**FIGURE 25D. ClustalW alignment of Drosophila protein with GadFly Accession Number CG11753; referred to as 'dCG11753' and human (hCG11753) and mouse (mCG11753) homologs. The sequences are shown in the one letter code; shaded residues match exactly.**

```

dCG11753 M K G G T F R N T Q W D P T L L S S Q I V S M Q F C Y Y F T 30
hCG11753 M - A G Q F R S Y V W D P L L I L S Q I V L M Q T V Y Y G S 29
mCG11753 M - A G Q F R S Y V W D P L L I L S Q I V L M Q T V Y Y G S 29

dCG11753 L G L L V F Y A - N K L S G D N Y S L D H L F E Y H E I H I Y 60
hCG11753 L G L W L A L Y D G L V R S S P S L D Q M F D A E I L G F S 59
mCG11753 L G L W L A L Y D A L V R S S P S L D Q M F D A E I L G F S 59

dCG11753 D M G G R L V I C A E V L N A F L A S L A L W C I Y R R A K 90
hCG11753 T P P G R L S M M S F I L N A L T C A L G L L Y F I R R G K 89
mCG11753 T P P G R L S M M S F E V L N A L T C A L G L L Y F I R R G K 89

dCG11753 L C L D F S C T F H Y L H L L I C W W Y N R S F P A N A S W 120
hCG11753 Q C L D F T V T V H F F H L L G C W F Y S S R F P S A L T W 119
mCG11753 Q C L D F T V T V H F F H L L G C W L Y S S R F P S A L T W 119

dCG11753 W E L N Y I T G T I M C I G G E F L C L Q T E M K E I P V G 150
hCG11753 W L V Q A V C I A L M A V I G E Y L C M R T E L K E I G D R 149
mCG11753 W L V Q A Y C I A L M A V I G E Y L C M R T E L K E I P L S 149

dCG11753 Y A A L N - - - - Q K S D Y 160
hCG11753 N L L I R F F F G Y E F P L 163
mCG11753 S A - - - - - P K S N Y . 157

```

**FIGURE 26. HUMAN HOMOLOG OF CG7262****FIGURE 26A. tBLASTN SEARCH RESULT FOR CG7262**

Homology to human gene ref NM\_014669; protein ref NP\_055484.1

dbj|BAA07680.1| (D42085) KIAA0095 gene is related to *S.cerevisiae* NIC96 gene. [Homo sapiens]; Length = 819

Score = 482 bits (1227), Expect = e-134

Identities = 298/823 (36%), Positives = 456/823 (55%), Gaps = 44/823 (5%)

```

Query: 6   LLQRAQNLLNDAKAECLEPAVERTLQQVLRATTELHSRV---TQTGCKEIQAHILGSKG 62
          LLQ+A+ L  + +   ELP VER LQ++ +A  L SR   T      +++A +LLGS+G
Sbjct: 9   LLQQAELQAAETEGISELPHVERNQLQEIQQAGERLRSRTLRTSQTADVKASVLLGSRG 68

Query: 63  VDLPLKNQKLEALSARKTFEPLDVKADTDVRTFLKNERENAILSVIDTKKNISDSANKQ 122
          +D+  ++Q+LE+LSA  TFEPL+  DTD++ FLKNE++NA+LS IE+++K   A  +
Sbjct: 69  LDISHISQRLLESLSAATTFEPLFPVKDTDIQGFLKNEKDNALLSAIEESRKRTFGMAEY 128

Query: 123  RWASMNKVNNEKTRLLDALIAPSQNFIDLQRLPEPTIVNPLCQP-RSCLDPLELVYAE 181
          SM   W  + K R+L  L+A  ++ +D  +  EP+ ++ +  P RS LD +E+ YA++
Sbjct: 129  HRESMLVEWEQVKQRILHTLLASGEDALDFTQSEPSYISDVGPGRSSLDNIEMAYARQ 188

Query: 182  LRHYNELLLKSSHRPNLVQKFAHLSQSFGDCRLTDMWTLLACVTQISEPLRSDPIKSRQQ 241
          + YNE ++   +PNLV  A +++  D  ++DMWT++  +T  +   +D +K+R
Sbjct: 189  IYIYNEKIVNGHLQPNLVDLCASVAE-LDDKSISDMWTMVKQMTDVLTPATDALKNRSS 247

Query: 242  ---RPEFVTYAKSYLERRYRVFMC SQVGGSYAN-----NSYQLVVAYVNHFRFRAQQT 291
          R EFV  A +YLE+ Y+ +   V G+               +YQLV +++N +  A
Sbjct: 248  VEVRRMEFVRQALAYLEQSYKNTLVTVFGNLHQAQLGGVPGTYQLVRSFLNIKLPAPLP- 306

Query: 292  GLVD-TVREIPLWPLVYYGLRCGAVKVAVEFLREAGSSHDEFA---QLVADR NAGETNSK 347
          GL D  V   P+W L+YY +RCG +  A + +  A   EF   Q  +   +
Sbjct: 307  GLQDGEVEGHPVWALIYYCMRCGDLAASQVVNRAQHQLGEFKTWFEYMNKDRRLSPA 366

Query: 348  IENQLRLQYANKIRNSTDAYKKAVYCILLGCDVNEVHGEVAKTIDDFLWMRLAMI----- 402
          EN+LRL Y   +RN+TD YK+AVYCI+  CDV +  EVA  +D+LW++L  +
Sbjct: 367  TENKLRHLHYRRALRNNTDPYKRAVYCIIGRCDDVDNQSEVADKTEDYLWLKLNQVCFDDD 426

Query: 403  ---QPGDADNYGKLQSMILEQYGEKYFNARQQPYLYFETLALTGQFEAAIEFLARQDENR 459
          P D      + Q  +LE YGE +F  QQP+LYF+ L LT  QFEAA+ FL R +  R
Sbjct: 427  GTSSPQDRLTLSQFQKQLLEDYGESHTVNQQPFYLYFQVLFLLTAQFEAAVAFLFRMERLR 486

Query: 460  AHAVHMAIALFELGLLGSARSVSQPLLSIDIKDPQPLRRLNLTRLIRQYVQRFTDTSE 519
          HAVH+A+ LFEL LL  +  S  LLS +  DP  LRRLN  RL+  Y ++FE TD  E
Sbjct: 487  CHAVHVALVLFELKLLKSSGQSAQLLSHEPGDPCLRLRLNFRLLMLYTRKFESTDPRE 546

Query: 520  ALHYYYTLRCLKDSKGRNMFMACVCDLVVDSGVFDTTIFDLIFGKRQSSDQNDVNSGLFR 579
          AL Y+Y LR  KDS+G NMF+  CV +LV++S      FD+I GK ++      G+
Sbjct: 547  ALQYFYFLRDEKDSQGENMFLRCVSELVIES-----REFDMILGKLENDGSR--KPGVID 599

Query: 580  QFECPEFDTRTMAAQVGDELAALGNFEMSARLYEMAGKYNLAVKHICILMAQVVHLPTLG 639
          +F      DT+ +  +V      G FE +A+LY++A  +  ++ +  L++ VV  +
Sbjct: 600  KFTS----DTKPIINKVASVAENKGLFEEAAKLYDLAKNADKVLLELMNKLLSPVVPQISAP 656

Query: 640  GSLRERLGQDAQRFNQLLASDSIDVEPKMKSSFVLLQDLLIFFNFYHDGKFNAALDLLRQ 699
          S +ERL  A  +  +  I      +S+F LL DL+ FF+ YH G  + A D++ +
Sbjct: 657  QSNKERLKNMALSIAERYRAQGISANKFVDSTFYLLLDLITFFDEYHSGHIDRAFDIIER 716

Query: 700  TQLVPNTLDDVDVVLGNVKQLSGEVIKVLDPDIVAAMEIAQKQYKQLKAGSGSSLEQTQM 759

```

46/48

+LVP + V+ + + S E+ L +V++A M I Q+K+LK S SS + Q  
 Sbjct: 717 LKLVPLNQESVEERVAFAFRNFSDEIRHNLSEVLLATMNILFTQFKRLKGTSPSSSSSRPQR 776  
 Query: 760 -----QQLRQRAKALSNMAATMPYRLPNDTNKRLIELELDMH 796  
 QLR +A+ L A +PYR DTN RL+++E+ M+  
 Sbjct: 777 VIEDRDSQLRSQARTLITFAGMIPYRTSGDTNARLVQMEVLMN 819

**FIGURE 26B. Predicted coding sequence for the human homolog of CG7262 (2460 base pairs); (SEQ ID NO:36)**

atggatactgaggggtttggtgagctccttcagcaagctgaacagcttgcgtgctgagactgagggcatctcagagcttcc  
 ccatgtggaacggaacttacaggagatccagcaggcgaggagagcgctgcgttcccgtaccctaacacgcacgtcccagg  
 agacggcagatgtcaaggcgtcagtttctcctcggtctcggggacttgacatatcccatctcccagcgattggagagt  
 ctgagtgcagccaccacctttgagcctcttgagcctgtgaaggacactgacattcagggcttcttgaagaatgagaagga  
 caatgccctgctgtctgccatcgaagagctcccggaagaggaccttcggcatggctgaggagtaccatcgggagtcattgt  
 tggttgagtgaggagcaagtgaacagcgcaattctgcacacactgctggcatcaggagaagacgccccttgactttactcaa  
 gaaagcgagccaagctacatcagtgatgtgggacccctgggtcgaagctctcttgataacatcgagatggcctatgcgcg  
 gcaaatattatctataatgagaaaattgtaaatggacacctgcagcctaacctgggtggacctttgtgcttcgcgcgag  
 agctggatgataagagcatttccgacatgtggacatggtaaaacaaatgacagacgtgtgtgtgacacgggcaacggat  
 gccctgaagaaccgcagcagcgtggaagtgcgcagtgagttgtcaggcaggccttggcgtaccttgagcagagttataa  
 gaattacacccttgtgactgtctttggaattttgcatcaggccagctgggcggggtgcctgggacttaccaattggttc  
 gaagtttctgaacattaaactgccagctcccttgccctggactacaggatggagaggtggaaggccatcctgtgtgggag  
 ctaatttactactgcatgcgctgtggagacctgcttgcgccttcacaggtagttaatcgagccagcaccagctgggaga  
 gtttaaaacttggttccaggagtacatgaacagcaaggacagaagattgtcccagctacggaaaacaagctccggtgc  
 attaccgtagggccctcaggaaacaatacagatccctacaagcgggcccgtgtactgtatcattggcagatgtgacgtcacc  
 gacaaccagagtgaagtggcggaacaaactgaggattacctgtggctgaagtgaaccaagtgtgttttgacgcagatgg  
 caccagctccccacaagacaggtcactctctcacagttccagaagcagttgtgtggaagactatggcgagtcaccattta  
 cgggtgaaccagcaacccttctctacttccaagtctgttccctgacagcgcagtttgaagcagcagtttgccttcttttc  
 cgcagtgagcggctgcgcgtgcatgtgtccatgtagcactgggtgctgtttgagctgaagctgcttttaagtcctcttg  
 acagagtgtcagctcctcagccacgagcctgggtgacctccttgccttgcggcggtgaacttcgtgcggtcctcctatgc  
 tgtacacccgggaagtgttgagtcacggacccaaggaggccctccagttacttctatttctcagggatgagaaagatagt  
 caaggagaaaacatgtttctgcgctgtgtgagtgagcttgtgattgaagccgagagttcgatatgattccttgggaaact  
 agagaatgacggaagttagaaagcctggagtcataagataagtttactagtgcacaaaagcctattatcaacaaagtgtctt  
 ctgtggcagaaaataaaggactgtttgaaggaggcagcaaaagctgtatgaccttgccaagaatgtgacaaggtaactggag  
 ctgatgaacaaactgtctgagcctgtcgtccccagatcagtgccccgcaatccaacaaggagaggtgaagaacatggc  
 actctccattgcccgaacgggtatagggctcaagggaataagcgcaataaaattgtggactccacgttctatcttcttttg  
 acttgatcaccttttttgacgagtatcatagtgttcatattgatagagcttttgatatcattgagcgcttgaagctgggtg  
 cccctgaatcaggaaagtgtggaagagagagtggtgctgttccagaaatttcagtgatgaaatcaggcacaacctctcaga  
 agtgcttcttgcaccatgaacatcttgttcacacagtttaaggagctcaaggggacaagtcctcctcgtcatccaggc  
 ccgacgagtcagcaggacggactctcaactccgaagcgaagccgacactctgattaccttctgtggaatgatacca  
 taccgaacgtctggggacaccaatgacgaggtggtgcagatggaggtcctcatgaattaa

**FIGURE 26C. Predicted amino acid sequence for the human homolog of CG7262 (819 amino acids) (SEQ ID NO:37)**

1	mdtegfge11	qqaqlaaet	egiselphve	rnlqeiqqag	erlrsrtltr	tsqetadvka
61	svllgsrgld	ishisqrles	lsaattfep1	epvkdtidqg	flknekdnal	lsaieesrkr
121	tfgmaeeyhr	esmlveweqv	kqrilhtlla	sgedaldfiq	esepsyisdv	gppgrssldn
181	iemayarqi	iynekiyngh	lqpnlvdlca	svaelddksi	sdmwtmvkqm	tdvlltpatd
241	alknrssvev	rmeivrqala	yleqsyknyt	lvtvfgnlhq	aqlggvpgty	qlvrsflnik
301	lpaplpgld	geveghpwa	liyycmrcgd	llaasqvvr	aqhqlgefkt	wfqeymnskd
361	rrlspatenk	lrlhyrralr	nntdpykrav	yciigrctvt	dnqsevadkt	edylwlklnq
421	vcfdddgts	pqdr1t1sqf	qkqlledyge	shftvnnqpf	lyfqvlf1ta	qfeaavaf1f
481	rmerlrchav	hvalvlfelk	11lkssggsa	qllshepgdp	pclrrlnfvr	llmlytrkfe
541	stdprealqy	fyflrdekds	qgenmflrcv	selviesref	dmilgklend	gsrkpgvidk
601	ftsdtkpiin	kvasvaenk	lfeaaaklyd	laknadkvle	lmnk11spvv	pqisapgsnk
661	erlknmali	aeryraggis	ankfvdstfy	11ldlitffd	eyhsghidra	fdiierlklv
721	plnqesveer	vaafnfsde	irhnlsevl1	atmn11ftqf	kr1kgtspss	ssrpqrvi
781	rdsq1rsqar	tlitfagmip	yrtsgdtnar	lvqmevlmn		

47/48

**FIGURE 27. HUMAN HOMOLOG OF CG4291****FIGURE 27A. BLASTN SEARCH RESULTS**

**Homology to human gene ref|XP\_049375.1| WW domain-containing binding protein 4 FBP21**

/garid=G2JPRCT\_96H3LQJ /chrom=13 /contig=NT\_024560.3 /start=240639  
 /end=263360 /strand=minus WBP4: WW domain binding protein 4 (formin binding protein 21) (formin binding protein 21;FBP21)  
 Length = 3951

Score = 57.4 bits (136), Expect = 1e-07  
 Identities = 42/142 (29%), Positives = 65/142 (45%), Gaps = 9/142 (6%)  
 Frame = +1

Query: 26 SVAFHESGKRHKMNVAKRITD-----ISRNSEKSERERQKMDAEIRKMEEAAMKSYAQD 79  
 SV FHE GK HK NVAKRI++ I + S +E +K E ME AA+K+Y +D  
 Sbjct: 1015 SVEFHERGKHNHKNVAKRISEVVCLT\*IKQKSLDKAKEEEKASKEFAAMEAAALKAYQED 1194

Query: 80 VHSRG---DMTARSINTVMXXXXXXXXXXXXXXXXXXXXQVDPMRLEGLSDEEDQRRVA 136  
 + G ++ SI V + P E+++++  
 Sbjct: 1195 LKRLGLESEILEPSITPV-----TSTIPPTSTSNQQKEKKEKKKRK 1317

Query: 137 PGKVTSDAAVPEASLWVEGKSDEGHTYYWNV 167  
 P WVEG + EG+ YY+++  
 Sbjct: 1318 KD-----PSKGRWVEGITSEGYHYYYDL 1386

Score = 47.3 bits (110), Expect = 1e-04  
 Identities = 14/22 (63%), Positives = 20/22 (90%)  
 Frame = +1

Query: 3 EYWKSNERKFCDKCKWLSDNK 24  
 +YWKS +KFCD+CKCW++DN+  
 Sbjct: 589 DYWKSQPKKFCDYCKCWIADNR 654

Score = 32.8 bits (73), Expect = 2.5  
 Identities = 14/31 (45%), Positives = 20/31 (64%)  
 Frame = +2

Query: 139 KVTSDAAVPEASLWVEGKSDEGHTYYWNVKT 169  
 KV S ++WVEG S++G TYY+N +T  
 Sbjct: 1496 KVFSSY\*TAVKTVWVEGLSEDGFTYYNTET 1588

**FIGURE 27B. Predicted coding sequence for the human homolog protein (SEQ ID NO:38)**

>DTT09253019.1 NT\_024560.3:complement(241731..259418)

ATGCAGGCAGCTGCCCTGAAAGCATACCAAGAGGATTTGAAAAGACTTGGCTTAGAGTCAGAAATTTTGG  
 AGCCAAGCATAACACCAGTAACCAGCACTATCCCACCTACCTCGACATCAAATCAACAGAAAGAAAAGAA  
 AGAAAAGAAGAAAAGAAAAGATCCTTCAAAGGGCAGATGGGTAGAAGGCATAACCTCTGAGGGTTAC  
 CATTACTATTATGATCTTATCTCAGGAGCATCTCAGTGGGAGAAACCTGAAGGATTTCAAGGAGACTTAA  
 AAAAGACAGCAGTGAAGACCGTTTGGGTAGAAGGTTTAAGTGAAGATGGTTTTACCTATTACTATAATAC

AGAAACAGGAGCAGAATCCAGATGGGAGAAACCTGATGATTTCAATCCACACACTAGTGATCTGCCTTCT  
AGTAAGGTCAATGAAAATTTCACTTGGCACCCCTAGATGAATCCAAATCATCAGATTCGCATAGTGATTCTG  
ATGGGGAACAGGAAGCAGAAGAAGGAGGGGTCTCTACAGAGACAGAAAAGCCAAAAATAAAGTTTAAGGA  
AAAAAATAAAATAGTGATGGAGGAAGTGACCCAGAAACACAGAAAGAAAAAAGTATTCAGAAACAGAAT  
TCATTAGGTTCAAATGAAGAAAAATCGAAAACCTCTTAAGAAATCAAACCCATATGGAGAATGGCAAGAAA  
TTAAACAAGAGGTTGAGTCTCATGAGGAGGTAGATTTGGAACTTCCAAGCACTGAAAATGAGTATGTATC  
AACTTCAGAAGCTGATGGTGGCGGAGAACCCTAAAGTGGTATTTAAAGAAAAAACAGTCACTTCTCTTGGA  
GTTATGGCAGATGGAGTGGCCCCAGTCTTCAAAAAGAGAAGAACTGAAAATGGAAAATCTAGAAATTTAA  
GGCAACGAGGTGATGATCAATAG

**FIGURE 27C. Predicted amino acid sequence for the human homolog protein  
(SEQ ID NO:39)**

MEAAALKAYQEDLKRLGLESEILEPSITPVTSTIPPTSTSNQQKEKKEKKKKRKKDPSKGRWVEGITSEGY  
HYYYDLISGASQWEKPEGFQGDLLKKTAVKTVWVEGLSEDGFTYYYNTETGAESRWEKPDDFIPTSDLPS  
SKVNENSLGTLDESKSSDSHSDSDGEQAEEGGVSTETEKPKIKFKEKNKNSDGGSDPETQKEKSIQKQN  
SLGSNEEKSKTLKKSHPYGEWQEIQEVESHEEVDLELPSTENEYVSTSEADGGGEPKVVFKEKTVTSLG  
VMADGVAPVFKKRRTEHGKSRNLRQRGDDQ